

KONINKRIJK DER



NEDERLANDEN

Bureau voor de Industriële Eigendom 18 JAN 2001

WIPO

PCT



Hierbij wordt verklaard, dat in Nederland op 24 september 1999 onder nummer 1013140.
ten name van:

DUPHAR INTERNATIONAL RESEARCH B.V.

te Weesp

een aanvraag om octrooi werd ingediend voor:

"Novel human-G protein coupled receptor ",

en dat de hieraan gehechte stukken overeenstemmen met de oorspronkelijk ingediende stukken.

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Rijswijk, 28 juli 2000.

De Directeur van het Bureau voor de Industriële Eigendom,
voor deze,

Mw. I.W. Scheevelenbos-de Reus.

Abstract

The present invention relates to novel identified polynucleotides, polypeptides encoded by them and to the use of such polynucleotides and polypeptides, and to their production. More particularly, the polynucleotides and polypeptides of the present invention relate to the G-protein coupled receptor family, referred to as IGS4-family. The invention also relates to inhibiting or activating the action of such polynucleotides and polypeptides, to a vector containing said polynucleotides, a host cell containing such vector and transgenic animals where the IGS4-gene is either overexpressed, misexpressed, underexpressed or suppressed (knock-out animals).

The invention further relates to a method for screening compounds capable to act as an agonist or an antagonist of said G-protein coupled receptor family IGS4 and the use of IGS4 polypeptides and polynucleotides and agonists or antagonists to the IGS4 receptor family in the treatment of CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine, attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis, dyslipidemias; Obesity; emesis; IBS, IBD, GERD, conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV 2; pain; cancers; urinary retention, ulcers, asthma; allergies and benign prostatic hypertrophy, among others and diagnostic assays for such conditions.

28 SEP. 1999

Description

5 The present invention relates to novel identified polynucleotides, polypeptides encoded by them and to the use of such polynucleotides and polypeptides, and to their production. More particularly, the polynucleotides and polypeptides of the present invention relate to a G-protein coupled receptor (GPCR), hereinafter referred to as IGS4. IGS4 exists in two polymorphic forms, hereinafter referred to as IGS4A and IGS4B. The invention also relates to inhibiting or activating the action of such

10 polynucleotides and polypeptides, to a vector containing said polynucleotides, a host cell containing such vector and transgenic animals where the IGS4-gene is either overexpressed, misexpressed, underexpressed or suppressed (knock-out animals). The invention further relates to a method for screening compounds capable to act as an agonist or an antagonist of said G-protein coupled receptor IGS4.

BACKGROUND OF THE INVENTION

15 It is well established that many medically significant biological processes are mediated by proteins participating in signal transduction pathways that involve G-proteins and/or second messengers; e.g., cAMP (Lefkowitz, Nature, 1991, 351:353-354). Herein these proteins are referred to as proteins

20 participating in pathways with G-proteins. Some examples of these proteins include the GPC receptors, such as those for adrenergic agents and dopamine (Kobilka, B.K., et al., Proc. Natl. Acad. Sci., USA, 1987, 84:46-50; Kobilka, B.K., et al., Science, 1987, 238:650-656; Bunzow, J.R., et al., Nature, 1988, 336:783-787), G-proteins themselves, effector proteins, e.g., phospholipase C, adenylyl cyclase, and

25 phosphodiesterase, and actuator proteins, e.g., protein kinase A and protein kinase C (Simon, M.I., et al., Science, 1991, 252:802-8).

30 For example, in one form of signal transduction, upon hormone binding to a GPCR the receptor interacts with the heterotrimeric G-protein and induces the dissociation of GDP from the guanine nucleotide-binding site. At normal cellular concentrations of guanine nucleotides, GTP fills the site immediately. Binding of GTP to the α subunit of the G-protein causes the dissociation of the G-protein from the receptor and the dissociation of the G-protein into α and $\beta\gamma$ subunits. The GTP-carrying form then binds to activated adenylyl cyclase. Hydrolysis of GTP to GDP, catalyzed by the G-protein itself (α subunit possesses an intrinsic GTPase activity), returns the G-protein to its basal, inactive form.

35 GTPase activity of the α subunit is, in essence, an internal clock that controls an on/off switch. The GDP bound form of the α subunit has high affinity for $\beta\gamma$ and subsequent reassociation of α GDP with $\beta\gamma$ returns the system to the basal state. Thus the G-protein serves a dual role, as an intermediate that relays the signal from receptor to effector (in this example adenylyl cyclase), and as a clock that controls the duration of the signal.

The membrane bound superfamily of G-protein coupled receptors has been characterized as having seven putative transmembrane domains. The domains are believed to represent transmembrane α -helices connected by extracellular or cytoplasmic loops. G-protein coupled receptors include a wide range of biologically active receptors, such as hormone, viral, growth factor and neuroreceptors.

The G-protein coupled receptor family includes dopamine receptors which bind to neuroleptic drugs used for treating CNS disorders. Other examples of members of this family include, but are not limited to calcitonin, adrenergic, neuropeptideY, somastotatin, neurotensin, neurokinin, capsaicin, VIP, CGRP, CRF, CCK, bradykinin, galanin, motilin, nociceptin, endothelin, cAMP, adenosine, muscarinic, acetylcholine, serotonin, histamine, thrombin, kinin, follicle stimulating hormone, opsin, endothelial differentiation gene-1, rhodopsin, odorant, and cytomegalovirus receptors.

Most G-protein coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structures. The 7 transmembrane regions are designated as TM1, TM2, TM3, TM4, TM5, TM6 and TM7. The cytoplasmic loop which connects TM5 and TM6 may be a major component of the G-protein binding domain.

Most G-protein coupled receptors contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxy terminus. For several G-protein coupled receptors, such as the β -adrenoreceptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization.

Recently, it was discovered that certain GPCRs, like the calcitonin-receptor like receptor, might interact with small single pass membrane proteins called receptor activity modifying proteins (RAMP's). This interaction of the GPCR with a certain RAMP is determining which natural ligands have relevant affinity for the GPCR-RAMP combination and regulate the functional signaling activity of the complex (McLathie, L.M. et al., Nature (1998) 393:333-339).

For some receptors, the ligand binding sites of G-protein coupled receptors are believed to comprise hydrophilic sockets formed by several G-protein coupled receptor transmembrane domains, said sockets being surrounded by hydrophobic residues of the G-protein coupled receptors. The hydrophilic side of each G-protein coupled receptor transmembrane helix is postulated to face inward and form a polar ligand-binding site. TM3 has been implicated in several G-protein coupled receptors as having a ligand-binding site, such as the TM3 aspartate residue. TM5 serines, a TM6 asparagine and TM6 or TM7 phenylalanines or tyrosines are also implicated in ligand binding.

G-protein coupled receptors can be intracellularly coupled by heteromeric G-proteins to various intracellular enzymes, ion channels and transporters (see, Johnson et al., Endoc. Rev., 1989, 10:317-331). Different G-protein α -subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of G-protein coupled receptors has been identified as an important mechanism for the regulation of G-protein coupling of some G-protein coupled receptors. G-protein coupled receptors are found in numerous sites within a mammalian host.

Receptors - primarily the GPCR class - have led to more than half of the currently known drugs (Drews, Nature Biotechnology, 1996, 14: 1516). This indicates that these receptors have an established, proven history as therapeutic targets. Clearly there is a need for identification and characterization of further receptors which can play a role in preventing, ameliorating or correcting dysfunctions or diseases, including, but not limited to CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, anorexia, bulimia, stroke, addiction/ dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy.

SUMMARY OF THE INVENTION

In one aspect, the invention relates to IGS4 polypeptides (including the IGS4A and IGS4B polypeptide polymorphs) and recombinant materials and methods for their production. Another aspect of the invention relates to methods for using such IGS4 polypeptides and polynucleotides. Such uses include the treatment of CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD;

conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy, among others.

- 5 In still another aspect, the invention relates to methods to identify agonists and antagonists using the materials provided by the invention, and treating conditions associated with IGS4 imbalance with the identified compounds. Yet another aspect of the invention relates to diagnostic assays for detecting diseases associated with inappropriate IGS4 activity or levels. A further aspect of the invention relates to animal-based systems which act as models for disorders arising from aberrant expression or activity of
- 10 IGS4.

Table 1: IGS4A-DNA of SEQ ID NO: 1 and SEQ ID NO: 3

5' -

5 GGGCTCAGCTTGAAACAGAGCCTCGTACCAGGGGAGGCTCAGGCCTTGGATTTTAATGTCA
GGGATGGAAAACTTCAGAATGCTTCCTGGATCTACCAGCAGAACTAGAAGATCCATTC
CAGAAACACCTGAACAGCACCGAGGAGTATCTGGCCTTCCTCTGCGGACCTCGGCGCAGC
CACTTCTTCCTCCCCGTGTCTGTGGTGTATGTGCCAATTTTTGTGGTGGGGGTCATTGGC
AATGTCTCTGGTGTGCCTGGTGAATCTGCAGCACCAGGCTATGAAGACGCCCACCAACTAC
TACCTCTTCAGCCTGGCGGTCTCTGACCTCCTGGTCCTGCTCCTTGGAAATGCCCCCTGGAG
10 GTCTATGAGATGTGGCGCAACTACCCTTTCTTGTTTCGGGCCCCGTGGGCTGCTACTTCAAG
ACGGCCCCCTTTGAGACCGTGTGCTTCGCCTCCATCCTCAGCATCACCACCGTCAGCGTG
GAGCGCTACGTGGCCATCCTACACCCGTTCCGCGCCAAACTGCAGAGCACCCGGCGCCGG
GCCCTCAGGATCCTCGGCATCGTCTGGGGCTTCTCCGTGCTCTTCTCCCTGCCCAACACC
AGCATCCATGGCATCAAGTTCCTACTTCCCCAATGGGTCCCTGGTCCCAGGTTCGGCC
15 ACCTGTACGGTCATCAAGCCCATGTGGATCTACAATTTTCATCATCCAGGTCACCTCCTTC
CTATTCTACCTCCTCCCCATGACTGTCTCAGTGTCTCTACTACCTCATGGCACTCAGA
CTAAAGAAAGACAAATCTCTTGAGGCAGATGAAGGGAATGCAAATATTCAAAGACCCTGC
AGAAAATCAGTCAACAAGATGCTGTTTGTCTTGGTCTTAGTGTTTGCTATCTGTTGGGCC
CCGTTCCACATTGACCGACTCTTCTTCAGCTTTGTGGAGGAGTGGAGTGAATCCCTGGCT
20 GCTGTGTTCAACCTCGTCCATGTGGTGTGAGGTGTCTTCTTCTACCTGAGCTCAGCTGTC
AACCCCATTTATCTATAACCTACTGTCTCGCCGCTTCCAGGCAGCATTCAGAAATGTGATC
TCTTCTTTCCACAAACAGTGGCACTCCCAGCATGACCCACAGTTGCCACCTGCCCAGCGG
AACATCTTCCTGACAGAATGCCACTTTGTGGAGCTGACCGAAGATATAGGTCCCCAATTC
CCATGTCAGTCATCCATGCACAACCTCTCACCTCCCCAACAGCCCTCTCTAGTGAACAGATG
25 TCAAGAACAACCTATCAAAGCTTCCACTTTAACAAAACCTGAATTCTTTTCAGAGCTGACT
CTCCTCTATGCCTCAAACTTCAGAGAGGAACATCCCATATATGTATGCCTTCTCATATGA
TATTAGAGAGGTAGAATGGCTCTTACAACCTCATGTACCCATTGCTAGTTTTTTTTTTT
ATAAACGTGAAAACCTGAGAGTTAGATCTGGTTTCAAACCCCAAGACTGCCTGATTTTTAG
TTATCTTTCCACTATCCTAACTGCCTCATGCCCTTCACTAGTTTCATGCCAAGAACGTGA
30 CTGGAAAGGCATGGCACCTATACCTTGATTAATTTCCATTAATGGAAATGGTTCGTCTCTG
AGTCATCTACGTTCCGAGTCAGGCTGTCACTCCTACTA-3'

Table 2: IGS4B-DNA of SEQ ID NO: 5 and SEQ ID NO: 7

5' -
GGCTCAGCCTTGAAACAGAGCCTCGTACCAGGGGAGGCTCAGGCCTTGGATTCTTAATGTCA
5 GGGATGGAAAACTTCAGAATGCTTCCTGGATCTACCAGCAGAACTAGAAGATCCATTG
CAGAAACACCTGAACAGCACCGAGGAGTATCTGGCCTTCCTCTGCGGACCTCGGCGCAGC
CACTTCTTCCTCCCCGTGTCTGTGGTGTATGTGCCAATTTTTGTGGTGGGGGTCATTGGC
AATGTCTCTGGTGTGCCTGGTGATTCTGCAGCACCAAGCTATGAAGACGCCCACCAACTAC
TACCTCTTCAGCCTGGCGGTCTCTGACCTCCTGGTCCTGCTCCTTGGAAATGCCCCCTGGAG
10 GTCTATGAGATGTGGCGCAACTACCCTTTCTTGTTTCGGGCCCCGTGGGCTGCTACTTCAAG
ACGGCCCTCTTTGAGACCGTGTGCTTCGCCTCCATCCTCAGCATCACCACCGTCAGCGTG
GAGCGCTACGTGGCCATCCTACACCCGTTCCGCGCCAAACTGCAGAGCACCCGGCGCCGG
GCCCTCAGGATCCTCGGCATCGTCTGGGGCTTCTCCGTGCTCTTCTCCCTGCCCAACACC
AGCATCCATGGCATCAAGTTCACCTACTTCCCCAATGGGTCCCTGGTCCCAGGTTCCGGCC
15 ACCTGTACGGTCATCAAGCCCATGTGGATCTACAATTTTCATCATCCAGGTCACCTCCTTC
CTATTCTACCTCCTCCCCATGACTGTCTCAGTGTCTCTACTACCTCATGGCACTCAGA
CTAAAGAAAGACAAATCTCTTGAGGCAGATGAAGGGAATGCAAATATTCAAAGACCCTGC
AGAAAATCAGTCAACAAGATGCTGTTTGTCTTGGTCTTAGTGTGTTGCTATCTGTTGGGCC
CCGTTCCACATTGACCGACTCTTCTTCAGCTTTGTGGAGGAGTGGACTGAATCCCTGGCT
20 GCTGTGTTCAACCTCGTCCATGTGGTGTCTCAGGTGTCTTATTCTACCTGAGCTCAGCTGTC
AACCCCATTTATCTATAACCTACTGTCTCGCCGCTTCAGGCAGCATTCCAGAATGTGATC
TCTTCTTTCCACAAACAGTGGCACTCCCAGCATGACCCACAGTTGCCACCTGCCCAGCGG
AACATCTTCCTGACAGAATGCCACTTTGTGGAGCTGACCGAAGATATAGGTCCCCAATTC
CTATGTCAATCATCCGTGCACAACTCTCACCTCCCAACAGCCCTCTCTAGTGAACAGATG
25 TCAAGAACAACTATCAAAGCTTCCACTTTAACAAAACCTGAATTCTTTTCAAGAGCTGACT
CTCCTCTATGCCTCAAACTTCAGAGAGGAACATCCCATAATGTATGCCTTCTCATATGA
AATTAGAGAGGTAGAATGGCTCTTACAACCTCATGTACCCATTGCTAGTTTTTTTTTTTA
ATAAACGTGAAAACCTGAGAGTTAGATCTGGTTTCAAACCCCAAGACTGCCTGATTTTTAG
TTATCTTTCCACTATCCTAACTGCCTCATGCCCCTTCACTAGTTCATGCCAAGAACGTGA
30 CTGGAAAGGCATGGCACCTATACCTTGATTAATTTCCATTAATGGAAATGGTTCGTCCTG
AGTCATCTACGTTCCGAGTCAGGCTGTCACTCCTACTA-3'

Table 4: IGS4A-protein of SEQ ID NO: 2 and SEQ ID NO: 4 (without the three amino acids between brackets).

5 (MSG) MEKLQNASWIYQQKLEDPFQKHLNSTEEYLAFLCGPRRSHFFLPVSVVYVPIFVVG
 IGNVLVCLVILQHQAAMKTPTNYYLFS LAVSDLLVLLLGMPLEVYEMWRNYPFLFGPVGCY
 FKTALFETVCFASILSITTVSVERYVAILHPFRAKLQSTRRRALRILGIVWGFSVLFSLP
 NTSIHGIKFHYFPNGSLVPGSATCTVIKPMWIYNFIIQVTSFLFYLLPMTVISVLYYLMA
 LRLKKDKSLEADEGNANIQRPCRKSVNKMLFVLVLVFAICWAPFHIDRLFFSFVEEWSES
 10 LAAVFNLVHVSVGVFFYLSSAVNP I IYNLLSRRFQAAFQNVISSFHKQWHSQHDPQLPPA
 QRNIFLTECHFVELTEDIGPQFPCQSSMHNSHLPTALSSEQMSRTNYQSFHFNKT

Table 5: IGS4B-protein of SEQ ID NO: 6 and SEQ ID NO: 8 (without the three amino acids between brackets).

5 (MSG) MEKLQNASWIYQQKLEDPFQKHLNSTEEYLAFLCGPRRSHFFLPVSVVYVPIFVVG
 IGNVLVCLVILQHQAAMKTPTNYYLFS LAVSDLLVLLLGMPLEVYEMWRNYPFLFGPVGCY
 FKTALFETVCFASILSITTVSVERYVAILHPFRAKLQSTRRRALRILGIVWGFSVLFSLP
 NTSIHGIKFHYFPNGSLVPGSATCTVIKPMWIYNFIIQVTSFLFYLLPMTVISVLYYLMA
 20 LRLKKDKSLEADEGNANIQRPCRKSVNKMLFVLVLVFAICWAPFHIDRLFFSFVEEWTES
 LAAVFNLVHVSVGVLFYLSSAVNP I IYNLLSRRFQAAFQNVISSFHKQWHSQHDPQLPPA
 QRNIFLTECHFVELTEDIGPQFLCQSSVHNSHLPTALSSEQMSRTNYQSFHFNKT

25

DESCRIPTION OF THE INVENTION

Definitions

30 The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"IGS4" refers, among others, to a polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 4 (IGS4A) and SEQ ID NO: 6 or SEQ ID NO: 8 (IGS4B), or an allelic
 35 variant thereof.

"Receptor Activity" or "Biological Activity of the Receptor" refers to the metabolic or physiologic function of said IGS4 including similar activities or improved activities or these activities with decreased undesirable side effects. Also included are antigenic and immunogenic activities of said IGS4.

"IGS4-gene" refers to a polynucleotide comprising the nucleotide sequence set forth in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7 or allelic variants thereof and/or their complements.

5

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of a Fab or other immunoglobulin expression library.

10

"Isolated" means altered "by the hand of man" from the natural state and/or separated from the natural environment. Thus, if an "isolated" composition or substance that occurs in nature has been "isolated", it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

15

20

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is a mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" may also include triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications has been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

25

30

35

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well-described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide.

including the peptide backbone, the amino acid side-chains and the amino and carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol; cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter et al., "Analysis for protein modifications and nonprotein cofactors", *Meth. Enzymol.* (1990) 182:626-646 and Rattan et al., "Protein Synthesis: Posttranslational Modifications and Aging", *Ann. NY Acad. Sci.* (1992) 663:48-62.

"Variant" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, and deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity" is a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" per se does not

art-recognized meaning and can be calculated using published techniques. See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed.; Academic Press, New York, 1993; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, 1987; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H., and Lipton, D., SIAM J. Applied Math. (1988) 48:1073). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipton, D., SIAM J. Applied Math. (1988) 48:1073. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., Nucleic Acids Research (1984) 12(1):387), BLASTP, BLASTN, FASTA (Atschul, S.F. et al., J. Molec. Biol. (1990) 215:403).

As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence of SEQ ID NO: 1 is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence of SEQ ID NO: 1. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence, or in a number of nucleotides of up to 5% of the total nucleotides in the reference sequence there may be a combination of deletion, insertion and substitution. These mutations of the reference sequence may occur at the 5 or 3 terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference amino acid sequence of SEQ ID NO: 2 is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of SEQ ID NO: 2. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference

sequence. These alterations in the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

5

Polypeptides of the Invention

10 In one aspect, the present invention relates to IGS4 polypeptides (or IGS4 proteins). The IGS4 polypeptides include the polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8 and the polypeptide having the amino acid sequence encoded by the DNA insert contained in the deposit no. CBS102221 or deposit no. CBS102221, deposited on September 24, 1999 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands; as well as polypeptides comprising the amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8 and the polypeptide having the amino acid sequence encoded by the DNA insert contained in the deposit no. 15 CBS102221 or deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands and polypeptides comprising the amino acid sequence which have at least 80% identity to that of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8 and/or the polypeptide having the amino acid sequence encoded by the DNA insert contained in the deposit no. CBS102221 or deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands over its entire 20 length, and still more preferably at least 90% identity, and even still more preferably at least 95% identity to said amino acid sequences. Furthermore, those with at least 97%, in particular at least 99%, are highly preferred. Also included within IGS4 polypeptides are polypeptides having the amino acid sequence which have at least 80% identity to the polypeptide having the amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8 or the polypeptide having the amino acid 25 sequence encoded by the DNA insert contained in the deposit no. CBS102221 or deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands over its entire length, and still more preferably at least 90% identity, and even still more preferably at least 95% identity to SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8. Furthermore, those with at least 97%, in particular at least 99% are highly preferred. Preferably IGS4 polypeptides exhibit at least one 30 biological activity of the receptor.

The IGS4 polypeptides may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification 35 such as multiple histidine residues, or an additional sequence for stability during recombinant production.

Fragments of the IGS4 polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that entirely is the same as part, but not all, of the amino

acid sequence of the aforementioned IGS4 polypeptides. As with IGS4 polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20; 21-40, 41-60, 61-80, 81-100; and 101 to the end of IGS4 polypeptide. In this context "about" includes the particularly recited ranges larger or smaller by several, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of IGS4 polypeptides, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. An example of a truncated polypeptide according to the present invention is the polypeptide of SEQ ID NO: 10 and SEQ ID NO: 12, which is encoded by the polynucleotide of SEQ ID NO: 9 respectively SEQ ID NO: 11. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Other preferred fragments are biologically active fragments. Biologically active fragments are those that mediate receptor activity, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

Thus, the polypeptides of the invention include polypeptides having an amino acid sequence at least 80% identical to that of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8 and/or the polypeptide having the amino acid sequence encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands, or fragments thereof with at least 80% identity to the corresponding fragment. Preferably, all of these polypeptide fragments retain the biological activity of the receptor, including antigenic activity. Variants of the defined sequence and fragments also form part of the present invention. Preferred variants are those that vary from the referents by conservative amino acid substitutions -- i.e., those that substitute a residue with another of like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination.

Table 3: IGS4A-64-DNA of SEQ ID NO: 9 and SEQ ID NO: 11

5' -

5 GGGTCAGCTTGAAACAGAGCCTCGTACCAGGGGAGGCTCAGGCCTTGGATTTTAATGTCA
 CAGAAACACCTGAACAGCACCAGGAGTATCTGGCCTTCCTCTGCGGACCTCGGCGCAGC
 CACTTCTTCTCCCCGTGTCTGTGGTGTATGTGCCAATTTTGTGGTGGGGGTCATTGGC
 AATGTCCTGGTGTGCCTGGTGATTCTGCAGCACCAGGCTATGAAGACGCCCACTAC
 TACCTCTTCAGCCTGGCGGTCTCTGACCTCCTGGTCCTGCTCCTTGAATGCCCCCTGGAG
 10 GTCTATGAGATGTGGCGCAACTACCCTTTCTTGTTCCGGCCCCGTGGGCTGCTACTTCAAG
 ACGGCCCTCTTTGAGACCGTGTGCTTCGCCTCCATCCTCAGCATCACCACCGTCAGCGTG
 GAGCGCTACGTGGCCATCCTACACCGGTTCCGCGCCAACTGCAGAGCACC CGCGCCGG
 GCCCTCAGGATCCTCGGCATCGTCTGGGGCTTCTCCGTGCTCTTCTCCCTGCCCAACACC
 AGCATCCATGGCATCAAGTTCCTACTTCCCCAATGGGTCCCTGGTCCCAGGTTCCGGCC
 15 ACCTGTACGGTCATCAAGCCCATGTGGATCTACAATTTTCATCATCCAGGTCACCTCCTTC
 CTATTCTACCTCCTCCCCATGACTGTCTCAGTGTCTCTACTACCTCATGGCACTCAGA
 CTAAGAAAGACAAATCTCTTGAGGCAGATGAAGGAATGCAAATATTCAAAGACCCTGC
 AGAAAATCAGTCAACAAGATGCTGTCTTTGTGGAGGAGTGGAGTGAATCCCTGGCTGCTG
 TGTTC AACCTCGTCCATGTGGTGTGAGGTGTCTTCTTCTACCTGAGCTCAGCTGTCAACC
 20 CCATTATCTATAACCTACTGTCTCGCCGCTTCCAGGCAGCATTCCAGAATGTGATCTCTT
 CTTTCCACAAACAGTGGCACTCCCAGCATGACCCACAGTTGCCACCTGCCAGCGGAACA
 TCTTCTGACAGAAAGCCACTTTGTGGAGCTGACCGAAGATATAGGTCCCCAATTCCCAT
 GTCAGTCATCCATGCACAACTCTCACCTCCCAACAGCCCTCTCTAGTGAACAGATGTCAA
 GAACAACTATCAAAGCTTCCACTTTAACAACCTGAATTCTTTTCAGAGCTGACTCTCC
 25 TCTATGCCTCAAACTTCAGAGAGGAACATCCCATATATGTATGCCTTCTCATATGATATT
 AGAGAGGTAGAATGGCTCTTACAACCTCATGTACCCATTGCTAGTTTTTTTTTTAATAA
 ACGTGAAAACCTGAGAGTTAGATCTGGTTTCAAAACCCCAAGACTGCCTGATTTTGTATTAT
 CTTTCCACTATCCTAACTGCCTCATGCCCCCTCACTAGTTTCATGCCAAGAACGTGACTGG
 AAAGGCATGGCACCTATACCTTGATTAATTTCCATTAATGGAAATGGTTCGTCCTGAGTC
 30 ATCTACGTTCCGAGTCAGGCTGTCACTCCTACTA-3'

Table 6: IGS4A-64-protein of SEQ ID NO: 10 and SEQ ID NO: 12 (without three amino acids between brackets).

35 MSKMEKLQNASWIYQQKLEDPFQPHLNSTEEYLAFLCGPRRSHEFLPVSVVYVPIFVWV
 IGVNLVCLVILCHQAMKTPTHYLLFSLAYSDLLVLLGMPLEVYEMWRNYPFLFGPVGGY
 FHTALFETVCFASILSITTVSVERYVAILHFFFAKLQSTERRALRILGIWGFVSLFSLP
 NGRHSHKPHYFPNGLVPGSATCTVTKPMWLNFIQVTCFLFYLLPMTHVLYVYLMMA
 LKLVYKQLEALFNNANLQFTFHFNNHMLCLWFSTNNFWLLQSTCCMWQVGGCT

The IGS4 polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Methods for preparing such polypeptides are well known in the art.

5

Polynucleotides of the Invention

A further aspect of the invention relates to IGS4 polynucleotides. IGS4 polynucleotides include isolated polynucleotides which encode the IGS4 polypeptides (including IGS4A and IGS4B) and
10 fragments, and polynucleotides closely related thereto. More specifically, the IGS4 polynucleotide of the invention includes a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7 encoding a IGS4A polypeptide of SEQ ID NO: 2 or of SEQ ID NO: 4 and a IGS4B polypeptide of SEQ ID NO: 6 or of SEQ ID NO: 8 respectively, polynucleotides having the particular sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ
15 ID NO: 7 and polynucleotides which essentially correspond to the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands.

IGS4 polynucleotides further include a polynucleotide comprising a nucleotide sequence that
20 has at least 80% identity over its entire length to a nucleotide sequence encoding the IGS4 polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8, a polynucleotide comprising a nucleotide sequence that is at least 80% identical to that of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7 over its entire length and a polynucleotide which essentially correspond to the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau
25 voor Schimmelcultures at Baarn the Netherlands.

In this regard, polynucleotides with at least 90% identity are particularly preferred, and those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred and those with at least 98-99% are most highly preferred, with at least 99% being the most preferred. Also included
under IGS4 polynucleotides are a nucleotide sequence which has sufficient identity to a nucleotide
30 sequence contained in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7 or to the DNA insert contained in the deposit no. CBS102221 or in the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands to hybridize under conditions useable for amplification or for use as a probe or marker. The invention also provides polynucleotides which are complementary to such IGS4 polynucleotides.

35

IGS4 of the invention is structurally related to other proteins of the G-protein coupled receptor family, as shown by the results of BLAST searches in the public databases. The amino acid sequence of Table 1 (SEQ ID NO: 2) has about 46 % identity (using BLAST, Altschul S.F. et al. [1997], Nucleic Acids Res. 25:3389-3402) over most of its length (amino acid residues 21-338) with a human orphan G-

protein coupled receptor (Accession # O43664, Tan et al., Genomics 52(2):217-229 (1998)). There is 27 % homology (over amino acid residues 61-349) to the rat neurotensin 1 receptor (Accession # P20789 Tanaka K. et al, Neuron 4:847-854 (1990)). The nucleotide sequence of Table 1 (SEQ ID NO: 1) is 63 % identical to an orphan G-protein coupled receptor over nucleotide residues 120-864 (Accession # AF044600, corresponding with the protein sequence O43664). Furthermore, there is 53 % identity to the human growth hormone secretagogue receptor over residues 94-1137 (Howard A.D. et al, Science 273:974-977(1996)). Thus, IGS4 polypeptides and polynucleotides of the present invention are expected to have, inter alia, similar biological functions/properties to their homologous polypeptides and polynucleotides, and their utility is obvious to anyone skilled in the art.

Polynucleotides of the invention can be obtained from natural sources such as genomic DNA. In particular, degenerated PCR primers can be designed that encode conserved regions within a particular GPCR gene subfamily. PCR amplification reactions on genomic DNA or cDNA using the degenerate primers will result in the amplification of several members (both known and novel) of the gene family under consideration (the degenerated primers must be located within the same exon, when a genomic template is used). (Libert et al., Science, 1989, 244: 569-572). Polynucleotides of the invention can also be synthesized using well-known and commercially available techniques.

The nucleotide sequence encoding the IGS4 polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8 may be identical to the polypeptide encoding sequence contained in SEQ ID NO: 1 (nucleotide number 55 to 1299) or SEQ ID NO: 3 (nucleotide number 64 to 1299), or SEQ ID NO: 5 (nucleotide number 55 to 1299) or SEQ ID NO: 7 (nucleotide number 64 to 1299) respectively, or it may be a different nucleotide sequence, which as a result of the redundancy (degeneracy) of the genetic code might also show alterations compared to the polypeptide encoding sequence contained in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7, but also encodes the polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8, respectively.

When the polynucleotides of the invention are used for the recombinant production of the IGS4 polypeptide, the polynucleotide may include the coding sequence for the mature polypeptide or a fragment thereof, by itself; the coding sequence for the mature polypeptide or fragment in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz et al., Proc. Natl. Acad. Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further preferred embodiments are polynucleotides encoding IGS4 variants comprising the amino acid sequence of the IGS4 polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8 in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acid residues are substituted, deleted or added, in any combination.

5

The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 80%, and preferably at least 90%, and more preferably at least 95%, yet even more preferably 97, in particular at least 99% identity between the sequences.

10

Polynucleotides of the invention, which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7 or a fragment thereof, may be used as hybridization probes for cDNA and genomic DNA, to isolate full-length cDNAs and genomic clones encoding IGS4 and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than human) that have a high sequence similarity to the IGS4 gene. People skilled in the art are well aware of such hybridization techniques. Typically these nucleotide sequences are 80% identical, preferably 90% identical, more preferably 95% identical to that of the referent. The probes generally will comprise at least 15 nucleotides. Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 30 and 50 nucleotides.

15

20

One embodiment, to obtain a polynucleotide encoding the IGS4 polypeptide, including homologs and orthologs from species other than human, comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7 or a fragment thereof, and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to those of skill in the art. Stringent hybridization conditions are as defined above or alternatively conditions under overnight incubation at 42 °C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1xSSC at about 65°C.

25

30

The polynucleotides and polypeptides of the present invention may be used as research reagents and materials for discovery of treatments and diagnostics to animal and human disease.

35

Vectors, Host Cells, Expression

The present invention also relates to vectors which comprise a polynucleotide or polynucleotides of the present invention, and host cells which are genetically engineered with vectors of the invention and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be used to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., BASIC METHODS IN MOLECULAR BIOLOGY (1986) and Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) such as calcium phosphate transfection, DEAE-dextran mediated transfection, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, E. coli, Streptomyces and Bacillus subtilis cells; fungal cells, such as yeast cells and Aspergillus cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used. Such systems include, among others, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al., MOLECULAR CLONING, A LABORATORY MANUAL (supra).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the desired polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If the IGS4 polypeptide is to be expressed for use in screening assays, generally, it is preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. In case the affinity or functional activity of the IGS4 polypeptide is modified by receptor activity modifying proteins (RAMP), coexpression of the relevant RAMP most likely at the surface of the cell is preferred and often required. Also in this event harvesting of cells expressing the IGS4 polypeptide and the relevant RAMP prior to use in screening assays is required. If the IGS4 polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide; if produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

IGS4 polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well-known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention also relates to the use of IGS4 polynucleotides for use as diagnostic reagents. Detection of a mutated form of the IGS4 gene associated with a dysfunction will provide a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of IGS4. Also in this event co-expression of relevant receptor activity modifying proteins can be required to obtain diagnostic assays of desired quality. Individuals carrying mutations in the IGS4 gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled IGS4 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers et al., *Science* (1985) 230:1242. Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method. See Cotton et al., *Proc. Natl. Acad. Sci. USA* (1985) 85: 4397-4401. In another embodiment, an array of

oligonucleotide probes comprising the IGS4 nucleotide sequence or segments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability. (See for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy, through detection of mutation in the IGS4 gene by the methods described.

In addition, CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy, can be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of the IGS4 polypeptide or IGS4 mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as an IGS4, in a sample derived from a host are well known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

In another aspect, the present invention relates to a diagnostic kit for a disease or susceptibility to a disease, particularly CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy, which comprises:

- (a) an IGS4 polynucleotide, preferably the nucleotide sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7, or a fragment thereof;
- (b) a nucleotide sequence complementary to that of (a);
- (c) an IGS4 polypeptide, preferably the polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or of SEQ ID NO: 8, or a fragment thereof; or
- (d) an antibody to an IGS4 polypeptide, preferably to the polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or of SEQ ID NO: 8.
- (e) a RAMP polypeptide required for the relevant biological or antigenic properties of an IGS4 polypeptide

It will be appreciated that in any such kit, (a), (b), (c) (d) or (e) may comprise a substantial component.

Chromosome Assays

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

The differences in the DNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

5 **Antibodies**

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them if required together with relevant RAMP's, may also be used as immunogens to produce antibodies immunospecific for the IGS4 polypeptides. The term "immunospecific" means that the antibodies have
10 substantial greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against the IGS4 polypeptides may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a nonhuman, using
15 routine protocols. For preparation of monoclonal antibodies, any technique, which provides antibodies produced by continuous cell line cultures, may be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Naure (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., Immunology Today (1983) 4:72) and the EBV-hybridoma technique (Cole et al., MONOCLONAL ANTIBODIES AND CANCER THERAPY, pp. 77-96, Alan R. Liss, Inc.,
20 1985).

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

25 Antibodies against IGS4 polypeptides as such or against IGS4 polypeptide-RAMP complexes, may also be employed to treat CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as
30 Huntington's disease or Gilles de la Tourette's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis;
35 infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; myocers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy, among others.

Animals

Another aspect of the invention relates to a animal-based systems which act as models for disorders arising from aberrant expression or activity of IGS4. Animal based model systems may also be used to further characterize the activity of the IGS4 gene. Such assays may be utilized as part of screening strategies designed to identify compounds which are capable to treat IGS4 based disorders such as CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy. In this way the animal-based models may be used to identify pharmaceutical compounds, therapies and interventions which may be effective in treating disorders aberrant expression or activity of IGS4. In addition such animal models may be used to determine the LD₅₀ and the ED₅₀ in animal subjects. These data may be used to determine the *in vivo* efficacy of potential IGS4 disorder treatments.

Animal-based model systems of IGS4 based disorders, based on aberrant IGS4 expression or activity, may include both non-recombinant animals as well as recombinantly engineered transgenic animals.

Animal models for IGS4 disorders may include, for example, genetic models. Animal models exhibiting IGS4 based disorder-like symptoms may be engineered by utilizing, for example, IGS4 sequences such as those described, above, in conjunction with techniques for producing transgenic animals that are well known to persons skilled in the art. For example, IGS4 sequences may be introduced into, and overexpressed and/or misexpressed in, the genome of the animal of interest, or, if endogenous IGS4 sequences are present, they may either be overexpressed, misexpressed, or, alternatively, may be disrupted in order to underexpress or inactivate IGS4 gene expression.

In order to overexpress or misexpress a IGS4 gene sequence, the coding portion of the IGS4 gene sequence may be ligated to a regulatory sequence which is capable of driving high level gene expression or expression in a cell type in which the gene is not normally expressed in the animal type of interest. Such regulatory regions will be well known to those skilled in the art, and may be utilized in the absence of undue experimentation.

For underexpression of endogenous IGS4 gene sequence, such sequence may be isolated and engineered such that when reintroduced into the genome of the animal of interest, the endogenous IGS4 gene alleles will be inactivated, or "knocked-out". Preferably, the engineered IGS4 gene sequence is introduced via gene targeting such that the endogenous IGS4 sequence is disrupted upon integration of the engineered IGS4 gene sequence into the animal's genome. Gene targeting is discussed, below, in this section.

Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats, and non-human primates, e.g., baboons, squirrels, monkeys, and chimpanzees may be used to generate animal models of IGS4 related disorders.

Any technique known in the art may be used to introduce a IGS4 transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to pronuclear microinjection (Hoppe, P.C. and Wagner, T.E., 1989, U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (van der Putten et al., Proc. Natl. Acad. Sci., USA 82:6148-6152, 1985); gene targeting in embryonic stem cells (Thompson et al., Cell 56:313-321, 1989.); electroporation of embryos (Lo, Mol. Cell. Biol. 3:1803-1814, 1983); and sperm-mediated gene transfer (Lavitrano et al., Cell 57:717-723, 1989); etc. For a review of such techniques, see Gordon, Transgenic Animals, Intl. Rev. Cytol. 115:171-229, 1989, which is incorporated by reference herein in its entirety.

The present invention provides for transgenic animals that carry the IGS4 transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i.e., mosaic animals. (See, for example, techniques described by Jakobovits, Curr. Biol. 4:761-763, 1994) The transgene may be integrated as a single transgene or in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko, M. et al., Proc. Natl. Acad. Sci. USA 89:6232-6236, 1992).

The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

When it is desired that the IGS4 transgene be integrated into the chromosomal site of the endogenous IGS4 gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous IGS4 gene of interest (e.g., nucleotide sequences of the mouse IGS4 gene) are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of, the nucleotide sequence of the endogenous IGS4 gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene of interest in only that cell type, by following, for example, the teaching of Gu et al. (Gu, H. et al., Science 265:103-106, 1994). The

regulatory sequences required for such a cell-type specific inactivation depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant IGS4 gene and protein may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the IGS4 transgene in the tissues of the transgenic animals may also be assessed using techniques which include but are not limited to Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and RT-PCR. Samples of target gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the target gene transgene product of interest. The IGS4 transgenic animals that express IGS4 gene mRNA or IGS4 transgene peptide (detected immunocytochemically, using antibodies directed against target gene product epitopes) at easily detectable levels may then be further evaluated to identify those animals which display characteristic IGS4 based disorder symptoms. Once IGS4 transgenic founder animals are produced (i.e., those animals which express IGS4 proteins in cells or tissues of interest, and which, preferably, exhibit symptoms of IGS4 based disorders), they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound IGS4 transgenics that express the IGS4 transgene of interest at higher levels because of the effects of additive expression of each IGS4 transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the possible need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; breeding animals to different inbred genetic backgrounds so as to examine effects of modifying alleles on expression of the IGS4 transgene and the development of IGS4-like symptoms. one such approach is to cross the IGS4 transgenic founder animals with a wild type strain to produce an F1 generation that exhibits IGS4 related disorder-like symptoms, such as those described above. The F1 generation may then be inbred in order to develop a homozygous line, if it is found that homozygous target gene transgenic animals are viable.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with the IGS4 polypeptide, or a fragment thereof, if required together with a RAMP polypeptide, adequate to produce antibody and/or T cell immune response to protect said animal from CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and

dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome; anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; Obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy, among others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises delivering the IGS4 polypeptide via a vector directing expression of the IGS4 polynucleotide in vivo in order to induce such an immunological response to produce antibody to protect said animal from diseases

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to an IGS4 polypeptide wherein the composition comprises an IGS4 polypeptide or IGS4 gene. The vaccine formulation may further comprise a suitable carrier. Since the IGS4 polypeptide may be broken down in the stomach, it is preferably administered parenterally (including subcutaneous, intramuscular, intravenous, intradermal etc. injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Screening Assays

The IGS4 polypeptide of the present invention may be employed in a screening process for compounds which bind the receptor and which activate (agonists) or inhibit activation of (antagonists) the receptor polypeptide of the present invention. Thus, polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries; and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics.

IGS4 polypeptides are responsible for biological functions, including pathologies. Accordingly, it is desirable to find compounds and drugs which stimulate IGS4 on the one hand and which can inhibit the

function of IGS4 on the other hand. In general, agonists are employed for therapeutic and prophylactic purposes for such conditions as CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy. Antagonists may be employed for a variety of therapeutic and prophylactic purposes for such conditions as CNS disorders, including schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as obsessive compulsive disorder (OCD), post traumatic stress disorder (PTSD), phobia and panic, major depressive disorder, bipolar disorder, Parkinson's disease, general anxiety disorder, autism, delirium, Alzheimer disease/dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, anorexia, bulimia, stroke, addiction/dependency/craving, sleep disorder, epilepsy, migraine; attention deficit/hyperactivity disorder (ADHD); cardiovascular disease including heart failure, angina pectoris, arrhythmias, myocardial infarction, hypotension; hypertension, thrombosis; dyslipidemias; obesity; emesis; IBS; IBD; GERD; conditions of delayed gastric emptying, such as postoperative or diabetic gastroparesis; osteoporosis; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; urinary retention; ulcers; asthma; allergies and benign prostatic hypertrophy.

In general, such screening procedures involve producing appropriate cells, which express the receptor polypeptide of the present invention on the surface thereof and, if essential co-expression of RAMP's at the surface thereof. Such cells include cells from mammals, yeast, *Drosophila* or *E. coli*. Cells expressing the receptor (or cell membrane containing the expressed receptor) are then contacted with a test compound to observe binding, or stimulation or inhibition of a functional response.

One screening technique includes the use of cells which express the receptor of this invention (for example, transfected CHO cells) in a system which measures extracellular pH or intracellular calcium changes caused by receptor activation. In this technique, compounds may be contacted with cells expressing the receptor polypeptide of the present invention. A second messenger response, e.g., signal transduction, pH changes, or changes in calcium level, is then measured to determine whether the potential compound activates or inhibits the receptor.

Another method involves screening for receptor inhibitors by determining inhibition or simulation of receptor-mediated cAMP and/or adenylate cyclase accumulation. Such a method involves transfecting an eukaryotic cell with the receptor of this invention to express the receptor on the cell surface. The cell is then exposed to potential antagonists in the presence of the receptor of this invention. The amount of cAMP accumulation is then measured. If the potential antagonist binds the receptor, and thus inhibits receptor binding, the levels of receptor-mediated cAMP, or adenylate cyclase, activity will be reduced or increased.

Another method for detecting agonists or antagonists for the receptor of the present invention is the yeast-based technology as described in U.S. Patent 5,482,835, incorporated by reference herein.

The assays may simply test binding of a candidate compound wherein adherence to the cells bearing the receptor is detected by means of a label directly or indirectly associated with the candidate compound or in an assay involving competition with a labeled competitor. Further, these assays may test whether the candidate compound results in a signal generated by activation of the receptor, using detection systems appropriate to the cells bearing the receptor at their surfaces. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed.

Further, the assays may simply comprise the steps of mixing a candidate compound with a solution containing an IGS4 polypeptide to form a mixture, measuring the IGS4 activity in the mixture, and comparing the IGS4 activity of the mixture to a standard.

The IGS4 cDNA, protein and antibodies to the protein may also be used to configure assays for detecting the effect of added compounds on the production of IGS4 mRNA and protein in cells. For example, an ELISA may be constructed for measuring secreted or cell associated levels of IGS4 protein using monoclonal and polyclonal antibodies by standard methods known in the art, and this can be used to discover agents which may inhibit or enhance the production of IGS4 (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues. Standard methods for conducting screening assays are well known in the art.

Examples of potential IGS4 antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligand of the IGS4, e.g., a fragment of the ligand, or small molecules which bind to the receptor but do not elicit a response, so that the activity of the receptor is prevented.

Thus in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for IGS4 polypeptides; or compounds which decrease or enhance the production of IGS4 polypeptides, which comprises:

- (a) an IGS4 polypeptide, preferably that of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8;
- (b) a recombinant cell expressing an IGS4 polypeptide, preferably that of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8;
- 5 (c) a cell membrane expressing an IGS4 polypeptide, preferably that of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8; or
- (d) antibody to an IGS4 polypeptide, preferably that of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

Prophylactic and Therapeutic Methods

This invention provides methods of treating abnormal conditions related to both an excess of and insufficient amounts of IGS4 activity.

15 If the activity of IGS4 is in excess, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as hereinabove described along with a pharmaceutically acceptable carrier in an amount effective to inhibit activation by blocking binding of ligands to the IGS4, or by inhibiting interaction with a RAMP polypeptide or a second signal, and thereby
20 alleviating the abnormal condition.

In another approach, soluble forms of IGS4 polypeptides still capable of binding the ligand in competition with endogenous IGS4 may be administered. Typical embodiments of such competitors comprise fragments of the IGS4 polypeptide.

25 In still another approach, expression of the gene encoding endogenous IGS4 can be inhibited using expression-blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered. See, for example, O'Connor, J Neurochem (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, Florida USA (1988). Alternatively, oligonucleotides, which form triple helices with the gene, can be supplied. See, for example, Lee et al., Nucleic Acids Res (1979) 6:3073; Cooney et al., Science (1988) 241:456; Dervan et al, Science (1991) 251:1360. These oligomers can be administered per se or the relevant oligomers can be expressed in vivo.

35 For treating abnormal conditions related to an under-expression of IGS4 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates IGS4, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of IGS4 by the

relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells in vivo and expression of the polypeptide in vivo. For overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics, Strachan T. and Read A.P., BIOS Scientific Publishers Ltd (1996).

Formulation and Administration

Peptides, such as the soluble form of IGS4 polypeptides, and agonists and antagonist peptides or small molecules, may be formulated in combination with a suitable pharmaceutical carrier. Such formulations comprise a therapeutically effective amount of the polypeptide or compound, and a pharmaceutically acceptable carrier or excipient. Formulation should suit the mode of administration, and is well within the skill of the art. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

Preferred forms of systemic administration of the pharmaceutical compositions include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if properly formulated in enteric or encapsulated formulations, oral administration may also be possible.

The dosage range required depends on the choice of peptide, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 $\mu\text{g/kg}$ of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

EXAMPLE 1. THE CLONING OF CDNA ENCODING A NOVEL G PROTEIN-COUPLED RECEPTOR.

Example 1a. Homology PCR cloning of a genomic fragment encoding a novel G-protein coupled receptor (GPCR).

A PCR based homology cloning strategy was used to isolate partial genomic DNA sequences encoding novel G-protein coupled receptors. Forward (F22) and reverse (R44 and R46) degenerate PCR primers were designed in conserved areas of the neurotensin receptor gene family (Vita N. et al.[1993] *Febs Lett.* 317, 139-142; Vita N. et al. [1998] *Eur. J. Pharmacol.* 360, 265-272) within transmembrane domains 1 (TM1) and 3 (TM3) and at the boundary between TM3 and intracellular loop 2 (I2):

F22 (TM1):

5'-CTCATCTTCGCGGTGGGC(A or G)C(A,C,G or T)G(C or T)(A,C,G or T)GG-3'

(SEQ ID NO: 13)

R44 (TM3/I2):

5'-GGCCAGGCAGCGCTCCGCGCT(C or Inosine)A(A or G)(A,C,G or T)C(C or T)(A,C,G or T)GC(A,G or T)-3'

(SEQ ID NO: 14)

R46 (TM3):

5'-GAA(A or G)TA(A or G)TAGCC(A or G)CG(A or G)CAGCC(A or T)-3'

(SEQ ID NO: 15)

In order to suppress amplification of known members of the neurotensin receptor family, the 3' ultimate nucleotide position of primer R44 was chosen in such a way that it was not complementary to the corresponding position of both NTR1 and NTR2 cDNA. The primary PCR reaction was carried out in a 50 µl volume and contained 100ng human genomic DNA (Clontech), 6 µl GeneAmp™ 10 x PCR buffer II (100mM Tris-HCl pH 8.3; 500 mM KCl, Perkin Elmer), 3.6 µl 25 mM MgCl₂, 0.36 µl dNTPs (25mM of each dNTP), 1.5 units AmpliTac-Gold™ polymerase (Perkin Elmer) and 30 pmoles of each of the degenerated forward (F22) and reverse primer (R44). Reaction tubes were heated at 95 °C for 10 min

and then subjected to 35 cycles of denaturation (95°C, 1 min), annealing (55°C, 2 min) and extension (72°C, 3min). Finally reaction tubes were heated for 10 min at 72°C.

For the semi-nested PCR reaction 1 µl of a 1/50 dilution of the primary PCR reaction was used as a template using the degenerate forward and reverse primers F22 and R46 respectively. The semi-nested PCR reaction was carried out under the same conditions as the primary PCR reaction.

Semi-nested PCR reaction products were size fractionated on an agarose gel and stained with ethidium bromide. Although a fragment of ± 220 bp was expected, only a fragment of ± 120 bp was visible. This fragment was purified from gel using the Qiaex-II™ purification kit (Qiagen) and ligated into the pGEM-T plasmid according to the procedure recommended by the supplier (pGEM-T kit, Promega). The recombinant plasmids thus produced were used to transform competent E. coli SURE™ 2 bacteria (Stratagene). Transformed cells were plated on LB agar plates containing ampicillin (100 µg/ml), IPTG (0.5 mM) and X-gal (50 µg/ml). Plasmid DNA was purified from mini-cultures of individual colonies using a Qiagen-tip 20 miniprep kit (Qiagen). DNA sequencing reactions were carried out on the purified plasmid DNA with the ABI Prism™ BigDye™ Terminator Cycle Sequencing Ready Reaction kit (PE-ABI), using insert-flanking primers.

Table 7: overview of oligo primers used.

| | |
|---------------|--|
| SEQ ID NO: 13 | F22: 5'-CTCATCTTCGCGGTGGGC(A or G)C(A,C,G or T)G(C or T)(A,C,G or T)GG-3' |
| SEQ ID NO: 14 | R44: 5'-GGCCAGGCAGCGCTCCGCGCT(C or Inosine)A(A or G)(A,C,G or T)C(C or T)(A,C,G or T)GC(A,G or T)-3' |
| SEQ ID NO: 15 | R46: 5'-GAA(A or G)TA(A or G)TAGCC(A or G)CG(A or G)CAGCC(A or T)-3' |
| SEQ ID NO: 16 | AP1: 5'-CCATCCTAATACGACTCACTATAGGGC-3' |
| SEQ ID NO: 17 | AP2: 5'-ACTCACTATAGGGCTCGAGCGGC-3' |
| SEQ ID NO: 18 | IGS4R1: 5'GGATCCCAAATAAGAAAGGGTAGTTGC-3' |
| SEQ ID NO: 19 | IGS4R2: 5'AAAGGGTAGTTGCGCCACATCTCATAGAC-3' |
| SEQ ID NO: 20 | IGS4F5: 5'AGGTCTATGAGATGTGGCGCAACTACCCT-3' |
| SEQ ID NO: 21 | IGS4F6: 5'ATGTGGCGCAACTACCCTTTCTTATTTGGG-3' |
| SEQ ID NO: 22 | R74: 5'-CGGAAGTTGGCGGACACG(A or G)(A,C or G)(A or G)TT(A or G)TA-3' |
| SEQ ID NO: 23 | IGS4F7: 5'-GCTCAGCTTGAAACAGAGCCTCGTACC-3' |
| SEQ ID NO: 24 | IGS4F8: 5'-CCATGTGGATCTACAATTCATCATCC-3' |
| SEQ ID NO: 25 | IGS4F9: 5'-AAGACAAATCTCTTGAGGCAGATGAAGGG-3' |
| SEQ ID NO: 26 | IGS4F10: 5'-GATGCTGTTTGTCTTGGTCTTAGTGTTC-3' |
| SEQ ID NO: 27 | IGS4R5: 5'-GGATGATGAAATTGTAGATCCACATGGGC-3' |
| SEQ ID NO: 28 | IGS4R6: 5'-TGTGGAGAAGTCTCTCAAAGTGTGG-3' |
| SEQ ID NO: 29 | IGS4R7: 5'-TAGTAGGAGTGACAGCCTGACTCGGAACG-3' |
| SEQ ID NO: 30 | IGS4R8: 5'-AACGTAGATGACTCAGGACGAACCATTTCC-3' |
| SEQ ID NO: 31 | IGS4F11: 5'-TCGTACCAGGGGAGGCTCAGGC-3' |

Sequencing reaction products were purified via EtOH/NaOAc precipitation and analyzed on an ABI 377 automated sequencer.

Sequence analysis of the insert of clone HNT1552 showed that it potentially encoded part of a novel member of the GPCR family. We refer to this novel GPCR sequence as IGS4.

Example 1b. Cloning of cDNA fragments containing the complete IGS4 coding sequence.

The complete coding sequence of IGS4 cDNA was obtained via both RACE analysis (rapid amplification of cDNA ends) and RT-PCR amplification. 5'- and 3' RACE PCR reactions were performed on Marathon-Ready™ cDNA from human brain or testis (Clontech n° 7400-1 and 7414-1 respectively), using the adaptor primers 1 and 2 (AP1: SEQ ID NO: 16 ; AP2: SEQ ID NO: 17) provided with the Marathon™ cDNA amplification kit (Clontech K1802-1) and IGS4 specific primers. PCR RACE reactions were performed according to the instructions of the Marathon-Ready™ cDNA user manual provided by Clontech. RACE products were separated on agarose gel, visualized with ethidium bromide and blotted onto Hybond N+ membranes. Blots were prehybridized at 65°C for 2 h in modified Church buffer (0.5M phosphate, 7% SDS, 10 mM EDTA) and then hybridised overnight at 65°C in the same buffer containing 2×10^6 cpm / ml of a ^{32}P -labelled IGS4 cDNA probe. IGS4 cDNA probes were radiolabelled via random primed incorporation of [α - ^{32}P]dCTP to a specific activity of $> 10^9$ cpm/ μg using the Prime-It II kit™ (Stratagene) according to the instructions provided by the supplier. Hybridized blots were washed at high stringency (2 x 30 min at room temperature in 2 x SSC/ 0.1% SDS, followed by 2 washes of 40 min at 65°C in 0.1 x SSC, 0.1% SDS) and autoradiographed overnight. Hybridizing fragments were purified from a preparative gel, cloned into the pGEM-T vector and sequenced as described above.

An initial round of semi-nested 5' RACE analysis on human brain cDNA using the IGS4 specific primers IGS4R1 (SEQ ID NO: 18) and IGS4R2 (SEQ ID NO: 19)(designed on the DNA sequence of clone HNT1552) yielded clones HNT1886 and HNT1887 (Fig.1). These clones extended the IGS4 cDNA sequence upstream up to and beyond the putative start of translation codon. Likewise an initial round of 3' RACE analysis on human brain cDNA using IGS4 specific primers IGS4F5 (SEQ ID NO: 20) and IGS4F6 (SEQ ID NO: 21) yielded clones HNT1874-1878 and HNT1902-1903 (Fig.1). These clones extended the known IGS4 cDNA at the 3' end.

All sequences obtained at this point were assembled into a single contig which contained a long open reading frame, encoding part of a novel protein that was most similar to human orphan receptor FM-3 (Tan et al., Genomics 52, 223-229 [1998], GenBank accession n° AF044600 and AF044601). To investigate the RNA expression profile of IGS4, a Master Blot™ membrane (Clontech cat n° 7770-1) containing RNA from different human tissues was hybridized to the ^{32}P -labelled insert of clone HNT1903 under the conditions recommended by the supplier. The strongest hybridization was obtained with testis RNA whereas much weaker signals were obtained in prostate, stomach, spinal cord, hippocampus, medulla oblongata, thyroid gland, thymus, lung and trachea.

Since the contig sequence did not yet contain the complete IGS4 coding sequence we set up an

RT-PCR homology cloning experiment on human total brain RNA using specific primer IGS4F6 (SEQ ID NO: 21) and a degenerated primer (R74, SEQ ID NO: 22), which was designed in a conserved area (at the TM7/C-terminal intracellular part) of the GPCR subfamily composed of the neurotensin receptors 1 and 2, the growth hormone secretagogue receptor (Howard A.D. et al.[1996] Science 273, 974-977) and the orphan GPCR FM-3 and GPR38 (McKee K.K. et al.[1997] Genomics 46, 426-434). RT-PCR reactions were carried out in a 50 µl volume on 500 ng total RNA from human brain using the Titan™ One Tube RT-PCR System (Boehringer catalogue n° 1,888,382) according to the recommendations of the supplier. Briefly, RT-PCR conditions were as follows: reverse transcription for 45 min at 55°C; 2 min denaturation at 94°C, followed by a touch-down PCR reaction of 20 cycles (30 sec denaturation at 94°C, 30 sec annealing at 60°C [-0.25°C/cycle] and 2 min extension at 68°C) and an additional round of 30 PCR cycles (30 sec denaturation at 94°C, 30 sec annealing at 55°C and 3 min [+ 5 sec/cycle] extension at 68°C). This was concluded with an extra extension step of 7 min at 68°C. Reaction products were analyzed via Southern blotting using the radiolabelled insert of clone HNT1903. A fragment of ± 690 bp that hybridized to the probe was purified from the gel (QiaexII™, Qiagen) and cloned into the pGEM-T vector yielding clones HNT2210-2212. Sequence analysis of these clones allowed to extend the existing IGS4 cDNA contig in the 3' direction.

Since the extended IGS4 cDNA contig still did not yet contain a translational stop codon, additional IGS4 specific 3' RACE primers were designed (IGS4F7-10, SEQ ID NO: 23-26)). Nested or semi-nested 3' RACE reactions were carried out on Marathon Ready™ cDNA from human testis (Clontech 7414-1). IGS4 specific bands (as assessed via Southern blot analysis using an IGS4 specific probe) were cloned into pGEM-T. This yielded clones HNT2289-90 (AP1/IGS4F5->AP2/IGS4F9), HNT2293-2295 (AP1/IGS4F6->AP2/IGS4F9), HNT2296-2297 (AP1/IGS4F7->AP2/IGS4F9), HNT2308-2310 (AP1/IGS4F8->AP2/IGS4F10) HNT2253 (AP1/IGS4F7->AP1/IGS4F5). An additional 5' RACE PCR reaction carried out on testis Marathon Ready™ cDNA yielded clones HNT2279-2281 (AP2/IGS4R6->AP2/IGS4R5). (note: AP1/IGS4F5->AP2/IGS4F9 e.g. indicates that clones were generated from an IGS4 specific fragment obtained after the primary RACE PCR reaction [using primer pair AP1/IGS4F5] was nested with primer pair AP2/IGS4F9).

Sequence analysis of these clones allowed to extend the existing IGS4 cDNA contig further in the 3' direction although the end of the IGS4 coding sequence was not yet been reached. A computer-assisted homology search (Blastn; Altschul S.F. et al., Nucleic Acids Res. (1997) 25:3389-3402) of the IGS4 contig DNA sequence against the expressed sequence tag (EST) database (dbest) showed the presence of an EST sequence (accession n° N45474) which overlapped with the 3' end of the IGS4 contig (near 100 % identity in the overlap area). EST N45474 further extended the IGS4 DNA contig at the 3' end into a translational stop codon and into the 3' untranslated region (3'-UTR). In addition another set of ESTs was identified which all covered the 3'-UTR of the IGS4 mRNA (Fig.2). Additional IGS4 specific primers (IGS4R7-8, SEQ ID NO: 29-30)) were designed within the 3'-UTR of these ESTs. Primary PCR reactions were carried out on Marathon Ready™ cDNA from human testis using various combinations of the IGS4F7 (SEQ ID NO: 23), IGS4F11 (SEQ ID NO: 31) and IGS4R7-8 (SEQ ID NO: 29-30) primers. PCR tubes were heated for 2 min at 95°C and then subjected to 35 cycles of

denaturation (95°C, 30 sec), annealing (65°C, 30 sec) and extension (72°C, 1 min 30 sec). Finally the reactions tubes were heated at 72°C for 10 min. Nested PCR reactions were also carried out under the same conditions. DNA fragments of \pm 1630 bp were purified from gel and cloned into the pGEM-T vector. The following clones were obtained: HNT2311, HNT2312 and HNT2317 (IGS4F7/IGS4R7->IGS4F11/IGS4R8); HNT2313, HNT2324, HNT2326 and HNT2328 (IGS4F11->IGS4R8); HNT2314, HNT2315 and HNT2322 (IGS4F11->R7). Clone HNT2363 was obtained from a purified 1630 bp PCR fragment, that was amplified from human testis Marathon Ready™ cDNA using the IGS4F11/R7 primer pair under the following slightly modified conditions. After an initial denaturation of 2 min at 94°C, PCR tubes were subjected to 15 cycles of denaturation [15 sec, 94°C], annealing [30 sec, 65°C] and extension [2 min, 72°C] followed by another 20 cycles of denaturation [15 sec, 94°C], annealing [30 sec, 65°C] and extension [2 min, 72°C; +10sec/cycle]. There was a final extension step of 7 min at 72°C. Sequence analysis of these clones allowed to assemble an IGS4 cDNA consensus sequence (Fig.1). Close inspection of all clones showed that they actually were of 2 sequence types, which differed at 5 nucleotide positions. These variant sequences correspond to a polymorphism within the human population. We refer to these different cDNA types as IGS4ADNA (SEQ ID NO: 1 and SEQ ID NO: 3) and IGS4BDNA (SEQ ID NO: 5 and SEQ ID NO: 7). The consensus sequence contained a long open reading frame that contained two in-frame start codons (positions 55-57 (SEQ ID NO: 1 and SEQ ID NO: 5) and 64-66 (SEQ ID NO: 3 and SEQ ID NO: 7) in IGS4ADNA and IGS4BDNA), predicting a protein of either 415 (SEQ ID NO: 2 and SEQ ID NO: 6) or 412 (SEQ ID NO: 4 and SEQ ID NO: 8) amino acids, which showed good homology to GPCR proteins. Hydropathy analysis (Kyte J. et al.[1982] J. Mol. Biol. 157: 105-132; Klein P. et al.[1985] Biochim. Biophys. Acta 815:468-476) of the protein also indicated the presence of 7 transmembrane domains. Since the first ATG initiator codon is within a weak "Kozak" translation initiation context and the second one is in a strong Kozak context, it is likely that the IGS4A/B protein starts at the second methionine and is 412 amino acids long (Kozak M. [1999] Gene 234, 187-208). However some (or even exclusive) initiation at the first ATG cannot be excluded. Among the five polymorphic nucleotides, four (positions 947, 999, 1202 and 1216 in IGS4A/BDNA) resulted in a switch in the encoded amino acid residue, whereas the fifth (pos 1381 in IGS4A/BDNA) was within the 3'-UTR. The respective predicted protein sequences are referred to as IGS4APROT (SEQ ID NO: 2 and SEQ ID NO: 4) and IGS4BPROT (SEQ ID NO: 6 and SEQ ID NO: 8). (note 1: the sequence of IGS4APROT and IGS4BPROT in this document is represented as the longest possible (415 amino acids) sequence but it is understood that the actual protein might be 3 amino acids shorter at the amino-terminus; for this reason the first 3 amino acids of IGS4APROT and IGS4BPROT in Table 4 and 5 have been bracketed) (note 2: In this document IGS4 refers to the IGS4 sequence in general, irrespective of the particular allelic type). Homology searches of the IGS4 protein sequence against public domain protein databanks showed best homology to the human orphan GPCR FM-3 (accession no O43664, Tan C.P., et al. Genomics (1998) 52: 223-229; 46% identity in IGS4A amino acid residues 26-342).

Homology searches of DNA databanks with the IGS4 cDNA sequence yielded a number of entries which were also derived from the IGS4 gene locus (Fig.2 for overview):

- 10 EST sequence entries (accession nrs W61169, AI432384, W61131, X023570, F01358, F03770, Z38158, R40869, R37725, H11333), 2 STS (sequence tagged sites) (accession nrs G20615 and G05725) and one genomic sequence (accession nr AQ078563) were discovered which were all derived from the 3'-UTR of IGS4 cDNA.
- 5 • EST accession n° N45474 encoded the 3' end of the IGS4 coding sequence and part of the 3' UTR (cfr supra).
- A 'working draft' high throughput genomic sequence (accession nr AC008571, version AC008571.1, deposited 3 AUG 1999), which consisted of 42 unordered contigs assembled in an arbitrary order was discovered in which we detected the entire IGS4 cDNA sequence in 4 separate areas. These
10 areas most likely correspond to the different IGS4 exons as they were flanked by canonical splice donor and acceptor sequences. On the basis of this analysis the position of the different exons in the IGS4ADNA (or IGS4BDNA) sequence can be defined as follows: exon1 (1-780), exon 2 (781-865), exon 3 (866-991) and exon 4 (992-1658). The AC008571genomic sequence is of the IGS4A allelic type.
- 15 • 6 overlapping EST entries (accession nrs H11359, R13890, R13353, F07531, F05108, F05107) were discovered of which the assembled DNA sequence overlapped at its 3' end with IGS4 exon2 and the beginning of exon 3. However the DNA sequence upstream of exon 2 was completely different from IGS4 exon1. Probably these six EST's are derived from transcripts which originated from an alternative promoter.
- 20 • Finally 2 genomic sequence entries (accession nrs AQ019411 and AQ015065) were discovered which encoded exon 2.

Among the many IGS4 cDNA clones that we isolated in the different experiments described above, we also discovered a number of clones that contained a 64 bp deletion (pos 866-929 in IGS4ADNA)
25 (besides a number of clones derived from unspliced [or partially spliced] transcripts). So far we only discovered truncated transcripts of the polymorphic type A. We refer to this splice variant cDNA sequence as IGS4A-64DNA (SEQ ID NO: 9 and SEQ ID NO: 11). Since this deletion occurs exactly at the exon 2/exon 3 boundary and since the last 2 nucleotides of the deleted fragment are "AG", it is likely that this deletion represents an alternative splicing event in which the "AG" within exon 3 served as a
30 splice acceptor. The IGS4A reading frame encoded by the splice variant is frameshifted beyond the deletion point. The encoded (truncated) protein of 296 amino acids is referred to as IGS4A-64PROT (SEQ ID NO: 10 and SEQ ID NO: 12). Hydropathy analysis of the IGS4A-64PROT sequence shows that this protein only contains 5 transmembrane domains (corresponding to TM domains 1-5 of IGS4APROT). This truncated receptor might have physiological relevance.

35 The bacterial strain harboring plasmid HNT2322 (containing the IGS4ADNA insert) was recloned after replating on LB agar plates containing 100 µg ampicillin/ml and deposited both in the Innogenetics N.V. strain list (ICCG4320) and at the "Centraalbureau voor Schimmelculturen (CBS)" in Baarn, The Netherlands (accession n° CBS102221). Plasmid DNA was prepared from the recloned
40 isolate and the insert was resequenced and found to be identical to the IGS4ADNA sequence.

The bacterial strain harboring plasmid HNT2363 (containing the IGS4BDNA insert) was

recloned after replating on LB agar plates containing 100 µg ampicillin and deposited both in the Innogenetics N.V. strain list (ICCG4340) and at the "Centraalbureau voor Schimmelculturen (CBS)" in Baarn, The Netherlands (accession n° CBS102222). Plasmid DNA was prepared from the recloned isolate and the insert was resequenced and found to be identical to the IGS4BDNA sequence.

5

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

10

The figures show:

Fig.1 Schematic representation of the relative positions of the different cDNA clones that were isolated to generate the consensus IGS4 cDNA sequence. 5' and 3' RACE primers that were used are also indicated (IGS4R# and IGS4F# respectively) as well as the position of EST accession n° N45474. Primer IGS4R6 was located within intron 1. Some clones (e.g. HNT2311, HNT2312 and HNT2253) were only partially sequenced (only the part that was sequenced is indicated). CONSENSUS A and CONSENSUS B denote the consensus sequence of IGS4 allelic types A and B respectively. The nucleotide that was identified at each of the 4 polymorphic positions is indicated (shaded boxes) for each clone. "S" indicates a sequence ambiguity in clones HNT2211 and HNT2212 and means either "C" or "T". The coding area of IGS4A and IGS4B consensus sequences is indicated with "***". As there were some remaining sequence ambiguities in the 5' end of the consensus sequence, the IGS4ADNA and IGS4BDNA sequences have only been taken from position 86 until the end

25

Fig.2 Schematic representation of the relative positions of different DNA database entries compared to the IGS4 cDNA sequence. The IGS4 cDNA sequence is indicated with the boxes (the position of the IGS4 coding sequences is indicated with the filled boxes). The relative position of IGS4 exons 1-4 is indicated above the IGS4 cDNA sequence ("=="). The parts of the genomic sequence AC008571 that encode exons 1->4 are indicated with AC008571a->d respectively. The position of these fragments within the AC008571 sequence are: AC008571a (13129-13908 of the reverse complement of AC008571), AC008571b (51676-51760 of AC008571), AC008571c (79978-80103 of the reverse complement of AC008571) and AC008571d (83060-83728 of the reverse complement of AC008571). G05725 and G20615 are STS (sequence tagged sequence) entries whereas F05107, F05108, F07531, R13353, R13890, H11359, N45474, W61169, AI432384, W61131, AI023570, F01358, F03770, Z38158, R40869, R37725, H11333 are EST entries. The parts of genomic clones AQ019411 and AQ015065 that contain IGS4 exon 2 are indicated with "...". The 5' part of EST sequences F05107, F05108, F07531,

30

35

R13353, R13890 and [REDACTED] 1359 which is totally different from the [REDACTED] S4 cDNA sequence is indicated with "**". AQ078563 is a genomic clone.



SEQUENCE LISTING

<110> Duphar International Research BV

5 <120> Novel human G-protein coupled receptor

<130> DIR 566

<140>

10 <141>

<150> 31

<170> PatentIn Ver. 2.1

15

<210> 1

<211> 1658

<212> DNA

<213> Homo sapiens

20

<220>

<221> CDS

<222> (55)..(1299)

<223> IGS4A Long version

25

<400> 1

ggctcagctt gaaacagagc ctcgtaccag gggaggctca ggccttggat tttta atg 57
 Met
 1

30

tca ggg atg gaa aaa ctt cag aat gct tcc tgg atc tac cag cag aaa 105
 Ser Gly Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys
 5 10 15

35

cta gaa gat cca ttc cag aaa cac ctg aac agc acc gag gag tat ctg 153
 Leu Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu
 20 25 30

40

gcc ttc ctc tgc gga cct cgg cgc agc cac ttc ttc ctc ccc gtg tct 201
 Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser
 35 40 45

45

gtg gtg tat gtg cca att ttt gtg gtg ggg gtc att ggc aat gtc ctg 249
 Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu
 50 55 60 65

50

gtg tgc ctg gtg att ctg cag cac cag gct atg aag acg ccc acc aac 297
 Val Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr Asn
 70 75 80

tac tac ctc ttc agc ctg ggc gtc tct gac ctc ctg gtc ctg ctc ctt 345
 Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Leu
 85 90 95

55

| | | |
|----|---|------|
| | gga atg ccc ctg gag gtc tat gag atg tgg cgc aac tac cct ttc ttg | 393 |
| | Gly Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe Leu | |
| | 100 105 110 | |
| 5 | ttc ggg ccc gtg ggc tgc tac ttc aag acg gcc ctc ttt gag acc gtg | 441 |
| | Phe Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr Val | |
| | 115 120 125 | |
| 10 | tgc ttc gcc tcc atc ctc agc atc acc acc gtc agc gtg gag cgc tac | 489 |
| | Cys Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg Tyr | |
| | 130 135 140 145 | |
| 15 | gtg gcc atc cta cac ccg ttc cgc gcc aaa ctg cag agc acc cgg cgc | 537 |
| | Val Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg Arg | |
| | 150 155 160 | |
| 20 | cgg gcc ctc agg atc ctc ggc atc gtc tgg ggc ttc tcc gtg ctc ttc | 585 |
| | Arg Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu Phe | |
| | 165 170 175 | |
| 25 | aat ggg tcc ctg gtc cca ggt tgc gcc acc tgt acg gtc atc aag ccc | 681 |
| | Asn Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys Pro | |
| | 195 200 205 | |
| 30 | atg tgg atc tac aat ttc atc atc cag gtc acc tcc ttc cta ttc tac | 729 |
| | Met Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe Tyr | |
| | 210 215 220 225 | |
| 35 | ctc ctc ccc atg act gtc atc agt gtc ctc tac tac ctc atg gca ctc | 777 |
| | Leu Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala Leu | |
| | 230 235 240 | |
| 40 | aga cta aag aaa gac aaa tct ctt gag gca gat gaa ggg aat gca aat | 825 |
| | Arg Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala Asn | |
| | 245 250 255 | |
| 45 | att caa aga ccc tgc aga aaa tca gtc aac aag atg ctg ttt gtc ttg | 873 |
| | Ile Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val Leu | |
| | 260 265 270 | |
| 50 | gtc tta gtg ttt got atc tgt tgg gcc ccg ttc cac att gac cga ctc | 921 |
| | Val Leu Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg Leu | |
| | 275 280 285 | |
| 55 | ttc ttc agc ttt gtg gag gag tgg agt gaa tcc ctg gct got gtg ttc | 969 |
| | Phe Phe Ser Phe Val Glu Glu Trp Ser Glu Ser Leu Ala Ala Val Phe | |
| | 290 295 300 305 | |
| 60 | aac ctc gtc cat gtg gtc tca ggt gtc ttc ttc tac ctg agc tca got | 1017 |
| | Asn Leu Val His Val Val Ser Gly Val Phe Phe Tyr Leu Ser Ser Ala | |
| | 310 315 320 | |

gtc aac ccc att atc tat aac cta ctg tct cgc cgc ttc cag gca gca 1065
 Val Asn Pro Ile Ile Tyr Asn Leu Leu Ser Arg Arg Phe Gln Ala Ala
 325 330 335
 5 ttc cag aat gtg atc tct tct ttc cac aaa cag tgg cac tcc cag cat 1113
 Phe Gln Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln His
 340 345 350
 10 gac cca cag ttg cca cct gcc cag cgg aac atc ttc ctg aca gaa tgc 1161
 Asp Pro Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu Cys
 355 360 365
 15 cac ttt gtg gag ctg acc gaa gat ata ggt ccc caa ttc cca tgt cag 1209
 His Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Pro Cys Gln
 370 375 380 385
 20 tca tcc atg cac aac tct cac ctc cca aca gcc ctc tct agt gaa cag 1257
 Ser Ser Met His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu Gln
 390 395 400
 atg tca aga aca aac tat caa agc ttc cac ttt aac aaa acc 1299
 Met Ser Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr
 405 410 415
 25 tgaattcttt cagagctgac tctcctctat gcctcaaaac ttcagagagg aacatcccat 1359
 aatgtatgcc ttctcatatg atattagaga ggtagaatgg ctcttacaac tcatgtaccc 1419
 30 attgctagtt tttttttttt aataaacgtg aaaactgaga gttagatctg gtttcaaaac 1479
 ccaagactgc ctgattttta gttatctttc cactatccta actgcctcat gcccttcac 1539
 tagttcatgc caagaacgtg actggaaagg catggcacct ataccttgat taatttccat 1599
 35 taatggaaat ggttcgtcct gagtcatcta cgttccgagt caggctgtca ctctacta 1658
 <210> 2
 40 <211> 415
 <212> PRT
 <213> Homo sapiens
 <400> 2
 45 Met Ser Gly Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln
 1 5 10 15
 Lys Leu Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr
 20 25 30
 50 Leu Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val
 35 40 45
 55 Ser Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val
 50 55 60

| | | | | | |
|----|---|-----|-----|-----|-----|
| | Leu Val Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr | 65 | 70 | 75 | 80 |
| 5 | Asn Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu | 85 | 90 | 95 | |
| | Leu Gly Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe | 100 | 105 | 110 | |
| 10 | Leu Phe Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr | 115 | 120 | 125 | |
| | Val Cys Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg | 130 | 135 | 140 | |
| 15 | Tyr Val Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg | 145 | 150 | 155 | 160 |
| | Arg Arg Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu | 165 | 170 | 175 | |
| | Phe Ser Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe | 180 | 185 | 190 | |
| 25 | Pro Asn Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys | 195 | 200 | 205 | |
| | Pro Met Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe | 210 | 215 | 220 | |
| 30 | Tyr Leu Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala | 225 | 230 | 235 | 240 |
| | Leu Arg Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala | 245 | 250 | 255 | |
| | Asn Ile Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val | 260 | 265 | 270 | |
| 40 | Leu Val Leu Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg | 275 | 280 | 285 | |
| | Leu Phe Phe Ser Phe Val Glu Glu Trp Ser Glu Ser Leu Ala Ala Val | 290 | 295 | 300 | |
| 45 | Phe Asn Leu Val His Val Val Ser Gly Val Phe Phe Tyr Leu Ser Ser | 305 | 310 | 315 | 320 |
| | Ala Val Asn Pro Ile Ile Tyr Asn Leu Leu Ser Arg Arg Phe Gln Ala | 325 | 330 | 335 | |
| 50 | Ala Phe Gln Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln | 340 | 345 | 350 | |
| 55 | | | | | |

His Asp Pro Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu
355 360 365

5 Cys His Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Pro Cys
370 375 380

Gln Ser Ser Met His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu
385 390 395 400

10 Gln Met Ser Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr
405 410 415

15 <210> 3
<211> 1658
<212> DNA
<213> Homo sapiens

20 <220>
<221> CDS
<222> (64)..(1299)
<223> IGS4A short version

25 <400> 3
ggctcagctt gaaacagagc ctggtaccag gggagggtca ggccttggat tttaatgtca 60

30 ggg atg gaa aaa ctt cag aat gct tcc tgg atc tac cag cag aaa cta 108
Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys Leu
1 5 10 15

35 gaa gat cca ttc cag aaa cac ctg aac agc acc gag gag tat ctg gcc 156
Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu Ala
20 25 30

40 ttc ctc tgc gga cct cgg cgc agc cac ttc ttc ctc ccc gtg tct gtg 204
Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser Val
35 40 45

45 gtg tat gtg cca att ttt gtg gtg ggg gtc att ggc aat gtc ctg gtg 252
Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu Val
50 55 60

50 tgc ctg gtg att ctg cag cac cag gct atg aag acg ccc acc aac tac 300
Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr Asn Tyr
65 70 75

55 tac ctc ttc agc ctg ggc gtc tct gac ctc ctg gtc ctg ctc ctt gga 348
Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Leu Gly
80 85 90 95

60 atg ccc ctg gag gtc tat gag atg tgg agc aac tac cct ttc ttg ttc 396
Met Pro Leu Gln Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe Leu Phe
100 105 110

| | | |
|----|---|------|
| | ggg ccc gtg ggc tgc tac ttc aag acg gcc ctc ttt gag acc gtg tgc | 444 |
| | Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr Val Cys | |
| | 115 120 125 | |
| 5 | ttc gcc tcc atc ctc agc atc acc acc gtc agc gtg gag cgc tac gtg | 492 |
| | Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg Tyr Val | |
| | 130 135 140 | |
| 10 | gcc atc cta cac ccg ttc cgc gcc aaa ctg cag agc acc cgg cgc cgg | 540 |
| | Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg Arg Arg | |
| | 145 150 155 | |
| 15 | gcc ctc agg atc ctc ggc atc gtc tgg ggc ttc tcc gtg ctc ttc tcc | 588 |
| | Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu Phe Ser | |
| | 160 165 170 175 | |
| 20 | ctg ccc aac acc agc atc cat ggc atc aag ttc cac tac ttc ccc aat | 636 |
| | Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe Pro Asn | |
| | 180 185 190 | |
| 25 | ggg tcc ctg gtc cca ggt tgc gcc acc tgt acg gtc atc aag ccc atg | 684 |
| | Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys Pro Met | |
| | 195 200 205 | |
| 30 | tgg atc tac aat ttc atc atc cag gtc acc tcc ttc cta ttc tac ctc | 732 |
| | Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe Tyr Leu | |
| | 210 215 220 | |
| 35 | ctc ccc atg act gtc atc agt gtc ctc tac tac ctc atg gca ctc aga | 780 |
| | Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala Leu Arg | |
| | 225 230 235 | |
| 40 | cta aag aaa gac aaa tct ctt gag gca gat gaa ggg aat gca aat att | 828 |
| | Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala Asn Ile | |
| | 240 245 250 255 | |
| 45 | caa aga ccc tgc aga aaa tca gtc aac aag atg ctg ttt gtc ttg gtc | 876 |
| | Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val Leu Val | |
| | 260 265 270 | |
| 50 | tta gtg ttt gct atc tgc tgg gcc ccg ttc cac att gac cga ctc ttc | 924 |
| | Leu Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg Leu Phe | |
| | 275 280 285 | |
| 55 | ttc agc ttt gtg gag gag tgg agt gaa tcc ctg got got gtg ttc aac | 972 |
| | Phe Ser Phe Val Glu Glu Trp Ser Glu Ser Leu Ala Ala Val Phe Asn | |
| | 290 295 300 | |
| 60 | ccc gtc cat gtg gtg tca ggt gtc ttc ttc tac ctg agc tca gat gtc | 1020 |
| | Leu Val His Val Val Ser Gly Val Phe Phe Tyr Leu Ser Ser Ala Val | |
| | 305 310 315 | |

aac ccc att atc tat aac cta ctg tct cgc cgc ttc cag gca gca ttc 1068
 Asn Pro Ile Ile Tyr Asn Leu Leu Ser Arg Arg Phe Gln Ala Ala Phe
 320 325 330 335

5 cag aat gtg atc tct tct ttc cac aaa cag tgg cac tcc cag cat gac 1116
 Gln Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln His Asp
 340 345 350

10 cca cag ttg cca cct gcc cag cgg aac atc ttc ctg aca gaa tgc cac 1164
 Pro Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu Cys His
 355 360 365

15 ttt gtg gag ctg acc gaa gat ata ggt ccc caa ttc cca tgt cag tca 1212
 Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Pro Cys Gln Ser
 370 375 380

20 tcc atg cac aac tct cac ctg cca aca gcc ctg tct agt gaa cag atg 1260
 Ser Met His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu Gln Met
 385 390 395

25 tca aga aca aac tat caa agc ttc cac ttt aac aaa acc tgaattcttt 1309
 Ser Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr
 400 405 410

30 cagagctgac tctcctctat gcctcaaaac ttcagagagg aacatcccat aatgtatgcc 1369
 ttctcatatg atattagaga ggtagaatgg ctcttacaac tcatgtaccc attgctagtt 1429
 tttttttttt aataaacgtg aaaactgaga gttagatctg gtttcaaaac ccaagactgc 1489
 ctgattttta gttatctttc cactatacta actgcctcat gcccttcac tagttcatgc 1549
 caagaacgtg actggaaagg catggcacct ataccttgat taatttccat taatggaaat 1609

35 ggttcgtcct gagtcatcta cgttccgagt caggctgtca ctctacta 1658

40 <210> 4
 <211> 412
 <212> PRT
 <213> Homo sapiens

45 <400> 4
 Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys Leu Glu
 1 5 10 15
 Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu Ala Phe
 20 25 30

50 Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser Val Val
 35 40 45

55 Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu Val Cys
 50 55 60

| | | | | | | | | | | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | Leu | Val | Ile | Leu | Gln | His | Gln | Ala | Met | Lys | Thr | Pro | Thr | Asn | Tyr | Tyr | 65 | 70 | 75 | 80 |
| 5 | Leu | Phe | Ser | Leu | Ala | Val | Ser | Asp | Leu | Leu | Val | Leu | Leu | Leu | Gly | Met | 85 | 90 | 95 | |
| | Pro | Leu | Glu | Val | Tyr | Glu | Met | Trp | Arg | Asn | Tyr | Pro | Phe | Leu | Phe | Gly | 100 | 105 | 110 | |
| 10 | Pro | Val | Gly | Cys | Tyr | Phe | Lys | Thr | Ala | Leu | Phe | Glu | Thr | Val | Cys | Phe | 115 | 120 | 125 | |
| | Ala | Ser | Ile | Leu | Ser | Ile | Thr | Thr | Val | Ser | Val | Glu | Arg | Tyr | Val | Ala | 130 | 135 | 140 | |
| 15 | Ile | Leu | His | Pro | Phe | Arg | Ala | Lys | Leu | Gln | Ser | Thr | Arg | Arg | Arg | Ala | 145 | 150 | 155 | 160 |
| 20 | Leu | Arg | Ile | Leu | Gly | Ile | Val | Trp | Gly | Phe | Ser | Val | Leu | Phe | Ser | Leu | 165 | 170 | 175 | |
| | Pro | Asn | Thr | Ser | Ile | His | Gly | Ile | Lys | Phe | His | Tyr | Phe | Pro | Asn | Gly | 180 | 185 | 190 | |
| 25 | Ser | Leu | Val | Pro | Gly | Ser | Ala | Thr | Cys | Thr | Val | Ile | Lys | Pro | Met | Trp | 195 | 200 | 205 | |
| 30 | Ile | Tyr | Asn | Phe | Ile | Ile | Gln | Val | Thr | Ser | Phe | Leu | Phe | Tyr | Leu | Leu | 210 | 215 | 220 | |
| | Pro | Met | Thr | Val | Ile | Ser | Val | Leu | Tyr | Tyr | Leu | Met | Ala | Leu | Arg | Leu | 225 | 230 | 235 | 240 |
| 35 | Lys | Lys | Asp | Lys | Ser | Leu | Glu | Ala | Asp | Glu | Gly | Asn | Ala | Asn | Ile | Gln | 245 | 250 | 255 | |
| | Arg | Pro | Cys | Arg | Lys | Ser | Val | Asn | Lys | Met | Leu | Phe | Val | Leu | Val | Leu | 260 | 265 | 270 | |
| 40 | Val | Phe | Ala | Ile | Cys | Trp | Ala | Pro | Phe | His | Ile | Asp | Arg | Leu | Phe | Phe | 275 | 280 | 285 | |
| | Ser | Phe | Val | Glu | Glu | Trp | Ser | Glu | Ser | Leu | Ala | Ala | Val | Phe | Asn | Leu | 290 | 295 | 300 | |
| 45 | Val | His | Val | Val | Ser | Gly | Val | Phe | Phe | Tyr | Leu | Ser | Ser | Ala | Val | Asn | 305 | 310 | 315 | 320 |
| 50 | Pro | Ile | Ile | Tyr | Asn | Leu | Leu | Ser | Arg | Arg | Phe | Gln | Ala | Ala | Phe | Gln | 325 | 330 | 335 | |
| | Asn | Val | Ile | Ser | Ser | Phe | His | Lys | Gln | Trp | His | Ser | Gln | His | Asp | Pro | 340 | 345 | 350 | |
| 55 | | | | | | | | | | | | | | | | | | | | |

Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu Cys His Phe
 355 360 365

5 Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Pro Cys Gln Ser Ser
 370 375 380

Met His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu Gln Met Ser
 385 390 395 400

10 Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr
 405 410

15 <210> 5
 <211> 1658
 <212> DNA
 <213> Homo sapiens

20 <220>
 <221> CDS
 <222> (55)..(1299)
 <223> IGS4B long version

25 <400> 5
 ggctcagctt gaaacagagc ctcgtagcagg gggagggtca ggccttggat ttta atg 57
 Met
 1

30 tca ggg atg gaa aaa ctt cag aat gct tcc tgg atc tac cag cag aaa 105
 Ser Gly Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys
 5 10 15

35 cta gaa gat cca ttc cag aaa cac ctg aac agc acc gag gag tat ctg 153
 Leu Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu
 20 25 30

40 gcc ttc ctc tgc gga cct cgg cgc agc cac ttc ttc ctc ccc gtg tct 201
 Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser
 35 40 45

gtg gtg tat gtg cca att ttt gtg gtg ggg gtc att ggc aat gtc ctg 249
 Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu
 50 55 60 65

45 gtg tgc ctg gtg att ctg cag cac cag gct atg aag acg ccc acc aac 297
 Val Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr Asn
 70 75 80

50 tac tac ctc ttc agc ctg gcg gtc tct gac ctc ctg gtc ctg ctc ctt 345
 Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Leu
 85 90 95

55

| | | |
|----|---|------|
| | gga atg ccc ctg gag gtc tat gag atg tgg cgc aac tac cct ttc ttg | 393 |
| | Gly Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe Leu | |
| | 100 105 110 | |
| 5 | ttc ggg ccc gtg ggc tgc tac ttc aag acg gcc ctc ttt gag acc gtg | 441 |
| | Phe Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr Val | |
| | 115 120 125 | |
| 10 | tgc ttc gcc tcc atc ctc agc atc acc acc gtc agc gtg gag cgc tac | 489 |
| | Cys Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg Tyr | |
| | 130 135 140 145 | |
| 15 | gtg gcc atc cta cac ccg ttc cgc gcc aaa ctg cag agc acc cgg cgc | 537 |
| | Val Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg Arg | |
| | 150 155 160 | |
| 20 | cgg gcc ctc agg atc ctc ggc atc gtc tgg ggc ttc tcc gtg ctc ttc | 585 |
| | Arg Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu Phe | |
| | 165 170 175 | |
| 25 | ttc ctg ccc aac acc agc atc cat ggc atc aag ttc cac tac ttc ccc | 633 |
| | Ser Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe Pro | |
| | 180 185 190 | |
| 30 | aat ggg tcc ctg gtc cca ggt tgc gcc acc tgt acg gtc atc aag ccc | 681 |
| | Asn Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys Pro | |
| | 195 200 205 | |
| 35 | atg tgg atc tac aat ttc atc atc cag gtc acc tcc ttc cta ttc tac | 729 |
| | Met Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe Tyr | |
| | 210 215 220 225 | |
| 40 | ctc ctc ccc atg act gtc atc agt gtc ctc tac tac ctc atg gca ctc | 777 |
| | Leu Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala Leu | |
| | 230 235 240 | |
| 45 | aga cta aag aaa gac aaa tct ctt gag gca gat gaa ggg aat gca aat | 825 |
| | Arg Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala Asn | |
| | 245 250 255 | |
| 50 | att caa aga ccc tgc aga aaa tca gtc aac aag atg ctg ttt gtc ttg | 873 |
| | Ile Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val Leu | |
| | 260 265 270 | |
| 55 | gtc tta gtg ttt gct atc tgt tgg gcc ccg ttc cac att gac cga ctc | 921 |
| | Val Leu Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg Leu | |
| | 275 280 285 | |
| 60 | ttc ttc agc ttt gtg gag gag tgg act gaa tcc ctg gct gct gtg ttc | 969 |
| | Phe Phe Ser Phe Val Glu Glu Trp Thr Glu Ser Leu Ala Ala Val Phe | |
| | 290 295 300 305 | |
| 65 | aac ctc gtc cat ggc gtg tca ggt gtc tta ttc tac ctg agc tca gct | 1017 |
| | Asn Leu Val His Val Val Ser Gly Val Leu Phe Tyr Leu Ser Ser Ala | |
| | 310 315 320 | |

gtc aac ccc att atc tat aac cta ctg tct cgc cgc ttc cag gca gca 1065
 Val Asn Pro Ile Ile Tyr Asn Leu Leu Ser Arg Arg Phe Gln Ala Ala
 325 330 335
 5 ttc cag aat gtg atc tct tct ttc cac aaa cag tgg cac tcc cag cat 1113
 Phe Gln Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln His
 340 345 350
 10 gac cca cag ttg cca cct gcc cag cgg aac atc ttc ctg aca gaa tgc 1161
 Asp Pro Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu Cys
 355 360 365
 15 cac ttt gtg gag ctg acc gaa gat ata ggt ccc caa ttc cta tgt cag 1209
 His Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Leu Cys Gln
 370 375 380 385
 20 tca tcc gtg cac aac tct cac ctc cca aca gcc ctc tct agt gaa cag 1257
 Ser Ser Val His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu Gln
 390 395 400
 atg tca aga aca aac tat caa agc ttc cac ttt aac aaa acc 1299
 Met Ser Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr
 405 410 415
 25 tgaattcttt cagagctgac tctcctctat gcctcaaaac ttcagagagg aacatcccat 1359
 aatgtatgcc ttctcatatg aaattagaga ggtagaatgg ctcttacaac tcatgtaccc 1419
 30 attgctagtt tttttttttt aataaacgtg aaaactgaga gttagatctg gtttcaaaac 1479
 ccaagactgc ctgattttta gttatctttc cactatccta actgcctcat gcccttcac 1539
 tagttcatgc caagaacgtg actggaaagg catggcacct ataccttgat taatttccat 1599
 35 taatggaaat ggttcgtcct gagtcaccta cgttcagagt caggctgtca ctctacta 1658
 <210> 6
 40 <211> 415
 <212> PRT
 <213> Homo sapiens
 <400> 6
 45 Met Ser Gly Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln
 1 5 10 15
 Lys Leu Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr
 20 25 30
 50 Leu Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val
 35 40 45
 Ser Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val
 50 55 60

Leu Val Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr
 65 70 75 80

5 Asn Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu
 85 90 95

Leu Gly Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe
 100 105 110

10 Leu Phe Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr
 115 120 125

Val Cys Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg
 130 135 140

15 Tyr Val Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg
 145 150 155 160

20 Arg Arg Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu
 165 170 175

Phe Ser Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe
 180 185 190

25 Pro Asn Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys
 195 200 205

Pro Met Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe
 210 215 220

Tyr Leu Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala
 225 230 235 240

35 Leu Arg Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala
 245 250 255

Asn Ile Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val
 260 265 270

40 Leu Val Leu Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg
 275 280 285

Leu Phe Phe Ser Phe Val Glu Glu Trp Thr Glu Ser Leu Ala Ala Val
 290 295 300

Phe Asn Leu Val His Val Val Ser Gly Val Leu Phe Tyr Leu Ser Ser
 305 310 315 320

50 Ala Val Asn Pro Ile Ile Tyr Asn Leu Leu Ser Arg Arg Phe Gln Ala
 325 330 335

Ala Phe Gln Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln
 340 345 350

55

His Asp Pro Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu
355 360 365

Cys His Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Leu Cys
370 375 380

Gln Ser Ser Val His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu
385 390 395 400

10 Gln Met Ser Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr
405 410 415

15 <210> 7
<211> 1658
<212> DNA
<213> Homo sapiens

20 <220>
<221> CDS
<222> (64)..(1299)
<223> IGS4B short version

25 <400> 7
ggctcagctt gaaacagagc ctggtaccag gggagggtca ggccttggat tttaatgtca 60

ggg atg gaa aaa ctt cag aat gct tcc tgg atc tac cag cag aaa cta 108
Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys Leu
30 1 5 10 15

gaa gat cca ttc cag aaa cac ctg aac agc acc gag gag tat ctg gcc 156
Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu Ala
20 25 30

35 ttc ctc tgc gga cct cgg cgc agc cac ttc ttc ctc ccc gtg tct gtg 204
Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser Val
35 40 45

40 gtg tat gtg cca att ttt gtg gtg ggg gtc att ggc aat gtc ctg gtg 252
Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu Val
50 55 60

45 tgc ctg gtg att ctg cag cac cag gct atg aag acg ccc acc aac tac 300
Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr Asn Tyr
65 70 75

50 tac ctc ttc agc ctg gcg gtc tct gac ctc ctg gtc ctg ctc ctt gga 348
Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Leu Gly
80 85 90 95

atg ccc ctg gag gtc tat gag atg tgg cgc aac tac cct ttc ttg ttc 396
Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe Leu Phe
100 105 110

55

| | | |
|----|---|------|
| | ggg ccc gtg ggc tgc tac ttc aag acg gcc ctc ttt gag acc gtg tgc | 444 |
| | Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr Val Cys | |
| | 115 120 125 | |
| 5 | ttc gcc tcc atc ctc agc atc acc acc gtc agc gtg gag cgc tac gtg | 492 |
| | Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg Tyr Val | |
| | 130 135 140 | |
| 10 | gcc atc cta cac ccg ttc cgc gcc aaa ctg cag agc acc cgg cgc cgg | 540 |
| | Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg Arg Arg | |
| | 145 150 155 | |
| 15 | gcc ctc agg atc ctc ggc atc gtc tgg ggc ttc tcc gtg ctc ttc tcc | 588 |
| | Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu Phe Ser | |
| | 160 165 170 175 | |
| 20 | ctg ccc aac acc agc atc cat ggc atc aag ttc cac tac ttc ccc aat | 636 |
| | Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe Pro Asn | |
| | 180 185 190 | |
| | ggg tcc ctg gtc cca ggt tgc gcc acc tgt acg gtc atc aag ccc atg | 684 |
| | Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys Pro Met | |
| | 195 200 205 | |
| 25 | tgg atc tac aat ttc atc atc cag gtc acc tcc ttc cta ttc tac ctc | 732 |
| | Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe Tyr Leu | |
| | 210 215 220 | |
| 30 | ctc ccc atg act gtc atc agt gtc ctc tac tac ctc atg gca ctc aga | 780 |
| | Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala Leu Arg | |
| | 225 230 235 | |
| 35 | cta aag aaa gac aaa tct ctt gag gca gat gaa ggg aat gca aat att | 828 |
| | Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala Asn Ile | |
| | 240 245 250 255 | |
| 40 | caa aga ccc tgc aga aaa tca gtc aac aag atg ctg ttt gtc ttg gtc | 876 |
| | Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val Leu Val | |
| | 260 265 270 | |
| | tta gtg ttc gct atc tgt tgg gcc ccg ttc cac att gac cga ctc ttc | 924 |
| | Leu Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg Leu Phe | |
| | 275 280 285 | |
| 45 | ttc agc ttt gtg gag gag tgg act gaa tcc ctg gct gct gtg ttc aac | 972 |
| | Phe Ser Phe Val Glu Glu Trp Thr Glu Ser Leu Ala Ala Val Phe Asn | |
| | 290 295 300 | |
| 50 | ctc gtc cat gtg gtg tca ggt gtc tta ttc tac ctg agc tca gct gtc | 1020 |
| | Leu Val His Val Val Ser Gly Val Leu Phe Tyr Leu Ser Ser Ala Val | |
| | 305 310 315 | |
| 55 | aac ccc att atc tat aac cta ctg tct cgc cgc ttc cag gca gca ttc | 1068 |
| | Asn Pro Ile Ile Tyr Asn Leu Leu Ser Arg Phe Glu Ala Ala Phe | |
| | 320 325 330 335 | |

cag aat gtg atc tct tct ttc cac aaa cag tgg cac tcc cag cat gac 1116
 Gln Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln His Asp
 340 345 350

5

cca cag ttg cca cct gcc cag cgg aac atc ttc ctg aca gaa tgc cac 1164
 Pro Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu Cys His
 355 360 365

10

ttt gtg gag ctg acc gaa gat ata ggt ccc caa ttc cta tgt cag tca 1212
 Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Leu Cys Gln Ser
 370 375 380

15

tcc gtg cac aac tct cac ctc cca aca gcc ctc tct agt gaa cag atg 1260
 Ser Val His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu Gln Met
 385 390 395

20

tca aga aca aac tat caa agc ttc cac ttt aac aaa acc tgaattcttt 1309
 Ser Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr
 400 405 410

25

cagagctgac tctcctctat gctcctcaaac ttcagagagg aacatcccat aatgtatgcc 1369
 ttctcatatg aaattagaga ggtagaatgg ctcttacaac tcatgtaccc attgctagtt 1429
 tttttttttt aataaacgtg aaaactgaga gttagatctg gtttcaaac ccaagactgc 1489
 ctgattttta gttatctttc cactatccta actgcctcat gcccttcac tagttcatgc 1549

30

caagaacgtg actggaaagg catggcacct ataccttgat taatttccat taatggaaat 1609
 ggttcgtcct gagtcattcta cgttcctgagt caggctgtca ctctacta 1658

35

<210> 8
 <211> 412
 <212> PRT
 <213> Homo sapiens

40

<400> 8
 Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys Leu Glu
 1 5 10 15

45

Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu Ala Phe
 20 25 30

Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser Val Val
 35 40 45

50

Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu Val Cys
 50 55 60

55

Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr Asn Tyr Tyr
 65 70 75 80

Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Leu Gly Met
 85 90 95

5 Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe Leu Phe Gly
 100 105 110

Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr Val Cys Phe
 115 120 125

10 Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg Tyr Val Ala
 130 135 140

Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg Arg Arg Ala
 145 150 155 160

15 Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu Phe Ser Leu
 165 170 175

20 Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe Pro Asn Gly
 180 185 190

Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys Pro Met Trp
 195 200 205

25 Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe Tyr Leu Leu
 210 215 220

Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala Leu Arg Leu
 225 230 235 240

30 Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala Asn Ile Gln
 245 250 255

35 Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val Leu Val Leu
 260 265 270

Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg Leu Phe Phe
 275 280 285

40 Ser Phe Val Glu Glu Trp Thr Glu Ser Leu Ala Ala Val Phe Asn Leu
 290 295 300

Val His Val Val Ser Gly Val Leu Phe Tyr Leu Ser Ser Ala Val Asn
 305 310 315 320

45 Pro Ile Ile Tyr Asn Leu Leu Ser Arg Arg Phe Gln Ala Ala Phe Gln
 325 330 335

50 Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln His Asp Pro
 340 345 350

Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu Cys His Phe
 355 360 365

55

Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Leu Cys Gln Ser Ser
 370 375 380

5 Val His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu Gln Met Ser
 385 390 395 400

Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr
 405 410

10

<210> 9

<211> 1594

<212> DNA

15 <213> Homo sapiens

<220>

<221> CDS

<222> (55)..(942)

20 <223> IGS4A truncated DNA long version

<400> 9

ggctcagctt gaaacagagc ctcgtaccag gggaggctca ggccttggat tttta atg 57
 Met
 1

25

tca ggg atg gaa aaa ctt cag aat gct tcc tgg atc tac cag cag aaa 105
 Ser Gly Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys
 5 10 15

30

cta gaa gat cca ttc cag aaa cac ctg aac agc acc gag gag tat ctg 153
 Leu Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu
 20 25 30

35

gcc ttc ctc tgc gga cct cgg cgc agc cac ttc ttc ctc ccc gtg tct 201
 Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser
 35 40 45

40

gtg gtg tat gtg cca att ttt gtg gtg ggg gtc att ggc aat gtc ctg 249
 Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu
 50 55 60 65

45

gtg tgc ctg gtg att ctg cag cac cag gct atg aag acg ccc acc aac 297
 Val Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr Asn
 70 75 80

50

tac tac ctc ttc agc ctg gcg gtc tct gac ctc ctg gtc ctg ctc ctt 345
 Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Leu
 85 90 95

gga atg ccc ctg gag gtc tat gag atg tgg cgc aac tac cct ttc ttg 393
 Gly Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe Leu
 100 105 110

55

| | | |
|----|--|------|
| | ttc ggg ccc gtg ggc tgc tac ttc aag acg gcc ctc ttc gag acc gtg | 441 |
| | Phe Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr Val | |
| | 115 120 125 | |
| 5 | tgc ttc gcc tcc atc ctc agc atc acc acc gtc agc gtg gag cgc tac | 489 |
| | Cys Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg Tyr | |
| | 130 135 140 145 | |
| 10 | gtg gcc atc cta cac ccg ttc cgc gcc aaa ctg cag agc acc cgg cgc | 537 |
| | Val Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg Arg | |
| | 150 155 160 | |
| 15 | cgg gcc ctc agg atc ctc ggc atc gtc tgg ggc ttc tcc gtg ctc ttc | 585 |
| | Arg Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu Phe | |
| | 165 170 175 | |
| 20 | tcc ctg ccc aac acc agc atc cat ggc atc aag ttc cac tac ttc ccc | 633 |
| | Ser Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe Pro | |
| | 180 185 190 | |
| 25 | aat ggg tcc ctg gtc cca ggt tgc gcc acc tgt acg gtc atc aag ccc | 681 |
| | Asn Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys Pro | |
| | 195 200 205 | |
| 30 | atg tgg atc tac aat ttc atc atc cag gtc acc tcc ttc cta ttc tac | 729 |
| | Met Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe Tyr | |
| | 210 215 220 225 | |
| 35 | ctc ctc ccc atg act gtc atc agt gtc ctc tac tac ctc atg gca ctc | 777 |
| | Leu Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala Leu | |
| | 230 235 240 | |
| 40 | aga cta aag aaa gac aaa tct ctt gag gca gat gaa ggg aat gca aat | 825 |
| | Arg Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala Asn | |
| | 245 250 255 | |
| 45 | att caa aga ccc tgc aga aaa tca gtc aac aag atg ctg tct ttg tgg | 873 |
| | Ile Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Ser Leu Trp | |
| | 260 265 270 | |
| 50 | agg agt gga gtg aat ccc tgg ctg ctg tgt tca acc tgc tcc atg tgg | 921 |
| | Arg Ser Gly Val Asn Pro Trp Leu Leu Cys Ser Thr Ser Ser Met Trp | |
| | 275 280 285 | |
| 55 | tgt cag gtg tct tct tct acc tgagctcagc tgtcaacccc attatctata | 972 |
| | Cys Gln Val Ser Ser Ser Thr | |
| | 290 295 | |
| 60 | acctactgtc tggccgcttc caggcagcat tccagaatgt gatctcttct ttccacaaac | 1032 |
| 65 | agtggcactc ccagcatgac ccacagttgc cacttgcaca ggggaacata ttcttgacag | 1092 |
| 70 | aatgccaact tgtggagctg atcgaagata taggtcccca attcccattgt cagtcatcca | 1152 |
| 75 | tgcacaaact tcaactccca acagccctct ctagtgaaca gatgtcaaga acaaaactat | 1212 |

aaagcttcca ctttaacaaa aactgaattc ttccagagct gactctcctc tatgcctcaa 1272
 aacttcagag aggaacatcc cataatgtat gccctctcat atgatattag agaggtagaa 1332
 5 tggctcttac aactcatgta cccattgcta gttttttttt ttttaataaac gtgaaaactg 1392
 agagtttagat ctgggtttcaa aacccaagac tgccctgattt ttagttatct ttccactatc 1452
 10 ctaactgcct catgccccct cactagttca tgccaagaac gtgactggaa aggcatggca 1512
 cctatacctt gattaatttc cattaatgga aatgggtcgt cctgagtcac ctacgttccg 1572
 agtcaggctg tcaactcctac ta 1594
 15

<210> 10

<211> 296

<212> PRT

20 <213> Homo sapiens

<400> 10

Met Ser Gly Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln
 1 5 10 15

25 Lys Leu Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr
 20 25 30

30 Leu Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val
 35 40 45

Ser Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val
 50 55 60

35 Leu Val Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr
 65 70 75 80

Asn Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu
 85 90 95

40 Leu Gly Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe
 100 105 110

45 Leu Phe Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr
 115 120 125

Val Cys Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg
 130 135 140

50 Tyr Val Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg
 145 150 155 160

Arg Arg Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu
 165 170 175

55

Phe Ser Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe
 180 185 190
 5 Pro Asn Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys
 195 200 205
 Pro Met Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe
 210 215 220
 10 Tyr Leu Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala
 225 230 235 240
 Leu Arg Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala
 245 250 255
 15 Asn Ile Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Ser Leu
 260 265 270
 Trp Arg Ser Gly Val Asn Pro Trp Leu Leu Cys Ser Thr Ser Ser Met
 275 280 285
 20 Trp Cys Gln Val Ser Ser Ser Thr
 290 295
 25
 <210> 11
 <211> 1594
 <212> DNA
 30 <213> Homo sapiens
 <220>
 <221> CDS
 <222> (64)...(942)
 35 <223> IGS4A-64 short version
 <400> 11
 ggctcagctt gaaacagagc ctcgtagcag gggaggctca ggccttggat tttaatgtca 60
 40 ggg atg gaa aaa ctt cag aat gct tcc tgg atc tac cag cag aaa cta 108
 Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys Leu
 1 5 10 15
 gaa gat cca ttc cag aaa cac ctg aac agc acc gag gag tat ctg gcc 156
 45 Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu Ala
 20 25 30
 ttc ctg tgc gga cct cgg cgc agc cac ttc ttc ctg ccc gtg tct gtg 204
 50 Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser Val
 35 40 45
 gtg tac gta cca att ttt ctg gta ggg gtc att ggc aat gta ctg gta 252
 Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu Val
 50 55 60

| | | |
|----|---|-----|
| | tgc ctg gtg att ctg cag cac cag gct atg aag acg ccc acc aac tac | 300 |
| | Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr Asn Tyr | |
| | 65 70 75 | |
| 5 | tac ctc ttc agc ctg gcg gtc tct gac ctc ctg gtc ctg ctc ctt gga | 348 |
| | Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Leu Gly | |
| | 80 85 90 95 | |
| 10 | atg ccc ctg gag gtc tat gag atg tgg cgc aac tac cct ttc ttg ttc | 396 |
| | Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe Leu Phe | |
| | 100 105 110 | |
| 15 | ggg ccc gtg ggc tgc cac ttc aag acg gcc ctc ttt gag acc gtg tgc | 444 |
| | Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr Val Cys | |
| | 115 120 125 | |
| 20 | ttc gcc tcc atc ctc agc atc acc acc gtc agc gtg gag cgc tac gtg | 492 |
| | Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg Tyr Val | |
| | 130 135 140 | |
| 25 | gcc atc cta cac ccg ttc cgc gcc aaa ctg cag agc acc cgg cgc cgg | 540 |
| | Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg Arg Arg | |
| | 145 150 155 | |
| 30 | ctg ccc aac acc agc atc cat ggc atc aag ttc cac tac ttc ccc aat | 636 |
| | Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe Pro Asn | |
| | 160 165 170 175 | |
| 35 | ggg tcc ctg gtc cca ggt tgg gcc acc tgt acg gtc atc aag ccc atg | 684 |
| | Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys Pro Met | |
| | 195 200 205 | |
| 40 | tgg atc tac aat ttc atc atc cag gtc acc tcc ttc cta ttc tac ctc | 732 |
| | Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe Tyr Leu | |
| | 210 215 220 | |
| 45 | ctc ccc atg act gtc atc agt gtc ctc tac tac ctc atg gca ctc aga | 780 |
| | Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala Leu Arg | |
| | 225 230 235 | |
| 50 | cta aag aaa gac aaa tct ctt gag gca gat gaa ggg aat gca aat att | 828 |
| | Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala Asn Ile | |
| | 240 245 250 255 | |
| 55 | caa aga ccc tgc aga aaa tca gtc aac aag atg ctg tct ttg tgg agg | 876 |
| | Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Ser Leu Trp Arg | |
| | 260 265 270 | |
| 60 | agt gga gag aat ccc tgg ctg ctg tgt tca acc tgg tcc atg tgg tgt | 924 |
| | Ser Gly Val Asn Pro Trp Leu Leu Cys Ser Thr Ser Ser Met Trp Cys | |
| | 275 280 285 | |

cag gtg tct tct tct acc tgagctcagc tgcacaacccc attatctata
Gln Val Ser Ser Ser Thr

971

290

5

acctactgtc tggcgcgttc caggcagcat tccagaatgt gatctcttct ttccacaaaac 1032

agtggcactc ccagcatgac ccacagttgc cacctgcccc goggaacatc ttccctgacag 1092

10

aatgccactt tgtggagctg accgaagata taggtcccca attcccatgt cagtcaccca 1152

tgcacaactc tcacctccca acagccctct ctagtgaaca gatgtcaaga acaaactatc 1212

15

aaagcttcca ctttaacaaa acctgaattc ttccagagct gactctcttc tatgcoctcaa 1272

aacttcagag aggaacatcc cataatgtat gccttctcat atgatattag agaggtagaa 1332

tggtctttac aactcatgta cccattgcta gttttttttt tttaataaac gtgaaaactg 1392

20

agagttagat ctggtttcaa aacccaagac tgctgattt ttagttatct ttccactatc 1452

ctaactgcct catgcccttc cactagttca tgccaagaac gtgactggaa aggcattggca 1512

cctatacctt gattaatttc cattaatgga aatgggtcgt cctgagtcac ctacgttcg 1572

25

agtcaggctg tcaactcctac ta

1594

30

<210> 12

<211> 293

<212> PRT

<213> Homo sapiens

<400> 12

35

Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln Lys Leu Glu

1

5

10

15

Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr Leu Ala Phe

20

25

30

40

Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val Ser Val Val

35

40

45

Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val Leu Val Cys

50

55

60

45

Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr Asn Tyr Tyr

65

70

75

80

50

Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu Leu Gly Met

85

90

95

Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe Leu Phe Gly

100

105

110

55

Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr Val Cys Phe
 115 120 125
 5 Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg Tyr Val Ala
 130 135 140
 Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg Arg Arg Ala
 145 150 155 160
 10 Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu Phe Ser Leu
 165 170 175
 Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe Pro Asn Gly
 180 185 190
 15 Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys Pro Met Trp
 195 200 205
 20 Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe Tyr Leu Leu
 210 215 220
 Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala Leu Arg Leu
 225 230 235 240
 25 Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala Asn Ile Gln
 245 250 255
 Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Ser Leu Trp Arg Ser
 260 265 270
 30 Gly Val Asn Pro Trp Leu Leu Cys Ser Thr Ser Ser Met Trp Cys Gln
 275 280 285
 35 Val Ser Ser Ser Thr
 290
 40 <210> 13
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Description of Artificial Sequence: Degenerated
 45 primers
 <220>
 <221> variation
 <222> (21)
 50 <223> A,C,G or T
 <220>
 <221> variation
 <222> (24)
 55 <223> A,C,G or T

<400> 13
ctcactctctcg cggctggggcgc agtggg

26

5 <210> 14
<211> 31
<212> DNA
<213> Artificial Sequence

10 <220>
<223> Description of Artificial Sequence: Degenerated
primers

15 <220>
<221> variation
<222> (22)
<223> C or Inosine

20 <220>
<221> variation
<222> (25)
<223> A,C,G or T

25 <220>
<221> variation
<222> (28)
<223> A,C,G or T

30 <400> 14
ggccaggcag cgctccgcgc tnarncyngc d

31

35 <210> 15
<211> 22
<212> DNA
<213> Artificial Sequence

40 <220>
<223> Description of Artificial Sequence: Degenerated
primers

<400> 15
gaartartag ccccgccagc cw

22

45 <210> 16
<211> 27
<212> DNA
<213> Artificial Sequence

50 <220>
<223> Description of Artificial Sequence: Primer

55 <400> 16
ccatctcaat acgactcaat atagggc

27

<210> 17
<211> 23
<212> DNA
<213> Artificial Sequence
5
<220>
<223> Description of Artificial Sequence: Primer

<400> 17
10 actcactata gggctcgagc ggc 23

<210> 18
<211> 27
15 <212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
20
<400> 18
ggatcccaaa taagaaaggg tagttgc 27

<210> 19
<211> 29
25 <212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
30

<400> 19
35 aaagggtagt tgcgccacat cccatagac 29

<210> 20
<211> 29
40 <212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 20
45 aggtctatga gatgtggcgc aactaccct 29

<210> 21
 <211> 30
 <212> DNA
 <213> Artificial Sequence

5

<220>
 <223> Description of Artificial Sequence: Primer

<400> 21
 atgtggcgca actacccttt cttatttggg

30

<210> 22
 <211> 26
 <212> DNA
 <213> Artificial Sequence

15

<220>
 <223> Description of Artificial Sequence : Degenerated
 primers

20

<400> 22
 cggaagtgg cgacacgrv rttrta

26

25

<210> 23
 <211> 27
 <212> DNA
 <213> Artificial Sequence

30

<220>
 <223> Description of Artificial Sequence: Primer

<400> 23
 gctcagcttg aaacagagcc tcgtacc

27

35

<210> 24
 <211> 27
 <212> DNA
 <213> Artificial Sequence

40

<220>
 <223> Description of Artificial Sequence: Primer

45

<400> 24
 ccattgtggat ctacaatttc atcatcc

27

5 <210> 25
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

10 <400> 25
aagacaaaatc tcttgaggca gatgaaggg 29

15 <210> 26
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
20

<400> 26
gatgctgttt gtcttggtct tagtgtttgc 30

25 <210> 27
<211> 29
<212> DNA
<213> Artificial Sequence

30 <220>
<223> Description of Artificial Sequence: Primer

35 <400> 27
ggatgatgaa attgtagatc cacatgggc 29

40 <210> 28
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

45 <400> 28
tgtggagaag tctctcaaag tgtgg 25

<210> 29
<211> 29
<212> DNA
<213> Artificial Sequence
5
<220>
<223> Description of Artificial Sequence: Primer

<400> 29
10 tagtaggagt gacagcctga ctcggaacg 29

<210> 30
<211> 30
15 <212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
20
<400> 30
aacgtagatg actcaggacg aaccatttcc 30

25 <210> 31
<211> 22
<212> DNA
<213> Artificial Sequence

30 <220>
<223> Description of Artificial Sequence: Primer

<400> 31
35 tcgtaccagg ggaggctcag gc 22

Claims

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

- 5 a) a nucleotide sequence encoding the IGS4 polypeptide according to SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8.
- b) a nucleotide sequence encoding the polypeptide encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the
- 10 Centraalbureau voor Schimmelcultures at Baarn the Netherlands, in particular a nucleotide sequence corresponding to the SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7.
- c) a nucleotide sequence having at least 80 % (preferably at least 90%) sequence identity over its entire length to the nucleotide sequence of (a) or (b);
- d) a nucleotide sequence which is complimentary to the nucleotide sequence of (a) or (b) or (c).
- 15

2. The polynucleotide of claim 1 wherein said polynucleotide comprises the nucleotide sequence contained in SEQ ID NO: 1 encoding the IGS4 polypeptide of SEQ ID NO: 2 or the nucleotide sequence contained in SEQ ID NO: 3 encoding the IGS4 polypeptide

20 of SEQ ID NO: 4 or the nucleotide sequence contained in SEQ ID NO: 5 encoding the IGS4 polypeptide of SEQ ID NO: 6 or the nucleotide sequence contained in SEQ ID NO: 7 encoding the IGS4 polypeptide of SEQ ID NO: 8.

3. The polynucleotide of claim 1 wherein said polynucleotide comprises a nucleotide sequence that is at least 80% identical to that of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7 or to the sequence of the DNA insert contained in the deposit

25 no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands over its entire length.

4. The polynucleotide of claim 3 which is the polynucleotide of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7 or the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor

30 Schimmelcultures at Baarn the Netherlands.

5. The polynucleotide of claim 1-4 which is DNA or RNA.

35

6. A DNA or RNA molecule comprising an expression system, wherein said expression system is capable of producing an IGS4 polypeptide comprising an amino acid sequence, which has at least 80% identity with the polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8 or with the polypeptide encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands, when said expression system is present in a compatible host cell.
7. A host cell comprising the expression system of claim 6.
8. A host cell according to claim 7 which is a yeast cell
9. A host cell according to claim 7 which is an animal cell
10. IGS4 receptor membrane preparation derived from a cell according to claim 7-9.
11. A process for producing an IGS4 polypeptide comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture.
12. A process for producing a cell which produces an IGS4 polypeptide comprising transforming or transfecting a host cell with the expression system of claim 6 such that the host cell, under appropriate culture conditions, produces an IGS4 polypeptide.
13. An IGS4 polypeptide comprising an amino acid sequence which is at least 80% identical to the amino acid sequence of SEQ ID NO: 2, SEQ NO: 4, SEQ NO: 6 or SEQ NO: 8 or to the polypeptide encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands over its entire length.
14. The polypeptide of claim 13 which comprises the amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ NO: 6 or SEQ NO: 8 or the amino acid sequence encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands.
15. An antibody immunospecific for the IGS4 polypeptide of claim 13.

16. A method for the treatment of a subject in need of enhanced activity or expression of IGS4 polypeptide of claim 13 comprising:

(a) administering to the subject a therapeutically effective amount of an agonist to said receptor; and/or

(b) providing to the subject an isolated polynucleotide comprising a nucleotide sequence that has at least 80% identity to a nucleotide sequence encoding the IGS4 polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ NO: 6 or SEQ NO: 8 or the polypeptide encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands over its entire length; or a nucleotide sequence complementary to one of said nucleotide sequences in a form so as to effect production of said receptor activity in vivo.

17. A method for the treatment of a subject having need to inhibit activity or expression of IGS4 polypeptide of claim 13 comprising:

(a) administering to the subject a therapeutically effective amount of an antagonist to said receptor; and/or

(b) administering to the subject a nucleic acid molecule that inhibits the expression of the nucleotide sequence encoding said receptor; and/or

(c) administering to the subject a therapeutically effective amount of a polypeptide that competes with said receptor for its ligand.

18. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of the IGS4 polypeptide of claim 13 in a subject comprising:

(a) determining the presence or absence of a mutation in the nucleotide sequence encoding said IGS4 polypeptide in the genome of said subject; and/or

(b) analyzing for the presence or amount of the IGS4 polypeptide expression in a sample derived from said subject.

19. A method for identifying agonists to the IGS4 polypeptide of claim 13 comprising:

(a) contacting a cell which produces a IGS4 polypeptide with a test compound; and

(b) determining whether the test compound effects a signal generated by activation of the IGS4 polypeptide.

20. An agonist identified by the method of claim 19.

21. The method for identifying antagonists to the IGS4 polypeptide of claim 13 comprising:
(a) contacting a cell which produces a IGS4 polypeptide with an agonist; and
(b) determining whether the signal generated by said agonist is diminished in the presence of a candidate compound.
- 5
22. An antagonist identified by the method of claim 21.
23. A recombinant host cell produced by a method of claim 12 or a membrane thereof expressing an IGS4 polypeptide.
- 10
24. A method of creating a genetically modified non-human animal comprising the steps of
(a) ligating the coding portion of a nucleic acid molecule, consisting essentially of a nucleic acid sequence encoding a protein having the amino acid sequence SEQ ID NO: 2, SEQ ID NO: 4, SEQ NO: 6 or SEQ NO: 8 or the amino acid sequence encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands or a biologically active portion of one of said sequences, to a regulatory sequence which is capable of driving high level gene expression or expression in a cell type in which the gene is not normally expressed in said animal; or
- 15
- 20 (b) isolation and engineering the coding portion of a nucleic acid molecule, consisting essentially of a nucleic acid sequence encoding a protein having the amino acid sequence SEQ ID NO: 2, SEQ ID NO: 4, SEQ NO: 6 or SEQ NO: 8 or the amino acid sequence encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands or a biologically active portion of one of said sequences, and reintroducing said sequence in the genome of said animal in such a way that the endogenous gene alleles, encoding a protein having the amino acid sequence SEQ ID NO: 2, SEQ ID NO: 4, SEQ NO: 6 or SEQ NO: 8 or the amino acid sequence encoded by the DNA insert contained in the deposit no. CBS102221 or the deposit no. CBS102222 at the Centraalbureau voor Schimmelcultures at Baarn the Netherlands or a biologically active portion of one of said sequences, are fully or partially inactivated.
- 25
- 30


```

F05107          *****
F05108          *****
F07531          *****
R13353          *****+--+
5  R13890          *****+--+
   H11359          *****+--+

AQ019411          +: +
AQ015065          +: +

10  AC008571a      +=====+
   AC008571b          +=+
   AC008571c          +=+
   AC008571d          +=====+
15  +-----1-----+2+3+-----4-----+
   IGS4DNA          [#####]
   N45474          +----->
   W61169          +----->
   G20615          <----+
20  AI432384          <-----+
   W61131          <-----+
   AI023570          <-----+
   F01358          <----+
   F03770          <----+
25  Z38158          <----+
   G05725          <----+
   R40869          <-----+
   R37725          <-----+
   H11333          <-----+
30  AQ078563          <-----+
                                <-----+
                                0         600       1200       1800       2400

```

Fig.2